

PATENT OFFICE JAPANESE GOVERNMENT

This is to certify that the annexed is a true copy of the following application as filed with the Office.

Date of Application : July 29, 2002
Application Number : Patent Application No. Heisei 2002-219103
Applicant (s) : FUJITSU LIMITED

February 7, 2001

Commissioner, Ohta Shinichiro
Patent Office

Certificate No. Toku 2003-3004888

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of:

Kimitaka MURASHITA

Application No.: TBA

Group Art Unit: TBA

Filed: July 1, 2003

Examiner: TBA

For: INFORMATION COLLECTING APPARATUS, METHOD, AND PROGRAM

**SUBMISSION OF CERTIFIED COPY OF PRIOR FOREIGN
APPLICATION IN ACCORDANCE
WITH THE REQUIREMENTS OF 37 C.F.R. § 1.55**

Commissioner for Patents
PO Box 1450
Alexandria, VA 22313-1450

Sir:

In accordance with the provisions of 37 C.F.R. § 1.55, the applicant(s) submit(s) herewith a certified copy of the following foreign application:

Japanese Patent Application No. 2002-219103

Filed: July 29, 2002

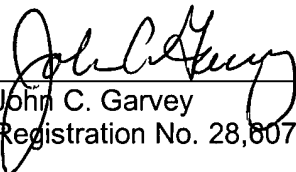
It is respectfully requested that the applicant be given the benefit of the foreign filing date as evidenced by the certified papers attached hereto, in accordance with the requirements of 35 U.S.C. § 119.

Respectfully submitted,

STAAS & HALSEY LLP

Date: 6-30-03

By: _____


John C. Garvey
Registration No. 28,807

1201 New York Ave, N.W., Suite 700
Washington, D.C. 20005
Telephone: (202) 434-1500
Facsimile: (202) 434-1501

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日
Date of Application:

2002年 7月29日

出 願 番 号
Application Number:

特願2002-219103

[ST.10/C]:

[JP2002-219103]

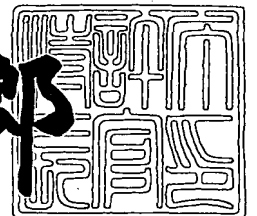
出 願 人
Applicant(s):

富士通株式会社

2003年 2月 7日

特 許 庁 長 官
Commissioner,
Japan Patent Office

太田信一郎



出証番号 出証特2003-3004888

【書類名】 特許願

【整理番号】 0251416

【提出日】 平成14年 7月29日

【あて先】 特許庁長官殿

【発明の名称】 情報収集装置、方法及びプログラム

【請求項の数】 5

【国際特許分類】 G06F 21/00

【発明者】

 【住所又は居所】 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

 【氏名】 村下 君孝

【特許出願人】

 【識別番号】 000005223

 【氏名又は名称】 富士通株式会社

【代理人】

 【識別番号】 100079359

 【弁理士】

 【氏名又は名称】 竹内 進

【手数料の表示】

 【予納台帳番号】 009287

 【納付金額】 21,000円

【提出物件の目録】

 【物件名】 明細書 1

 【物件名】 図面 1

 【物件名】 要約書 1

 【包括委任状番号】 9704823

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 情報収集装置、方法及びプログラム

【特許請求の範囲】

【請求項1】

ネットワークに接続するネットワーク接続部と、
ネットワークあるいは実社会上で発生したイベントの有無を検出するためのイベント収集先サイトを登録するイベント収集先サイト登録部と、
テキスト、画像、音声等のデータを含むドキュメントを収集するための情報収集先サイトを登録する情報収集先サイト登録部と、
前記登録されたイベント収集先サイトから情報を取得し、取得情報の更新の有無からイベント発生の有無を検出するイベント検出部と、
前記イベント検出部で検出された更新領域からキーワードを抽出するキーワード抽出部と、
前記キーワード抽出部で抽出されたキーワードを用いて前記登録された情報収集先サイト内のドキュメントを検索する情報検索部と、
前記情報検索部による検索結果をユーザに通知する情報通知部と、
を備えたことを特徴とする情報収集装置。

【請求項2】

イベント収集先サイト登録部により、ネットワークあるいは実社会上で発生したイベントの有無を検出するためのイベント収集先サイトを登録するイベント収集先サイト登録ステップと、
情報収集先サイト登録部により、テキスト、画像、音声等のデータを含むドキュメントを収集するための情報収集先サイトを登録する情報収集先サイト登録ステップと、
イベント検出部により、前記登録されたイベント収集先サイトから情報を取得し、取得情報の更新の有無からイベント発生の有無を検出するイベント検出ステ

ップと、

キーワード抽出部により、前記イベント検出ステップで検出された更新領域からキーワードを抽出するキーワード抽出ステップと、

情報検索部により、前記キーワード抽出部で抽出されたキーワードを用いて前記登録された情報収集先サイト内のドキュメントを検索する情報検索ステップと

情報通知部により、前記情報検索ステップの検索結果をユーザに通知する情報通知ステップと、

を備えたことを特徴とする情報収集方法。

【請求項3】

請求項2記載の情報収集方法に於いて、前記イベント検出ステップは、前記イベント収集先サイトをアクセスしてサイト内のドキュメントをダウンロードしてリファレンスとして保存し、その後、前記イベント収集先サイトからダウンロードしたドキュメントと前記リファレンスの比較により更新の有無からイベント発生の有無を検出することを特徴とする情報収集方法。

【請求項4】

請求項2記載の情報収集方法に於いて、更に、前記情報検索ステップにより前記情報収集先サイトから取得されたドキュメントをドキュメント保存部に保存するドキュメント保存ステップを設けたことを特徴とする情報収集方法。

【請求項5】

コンピュータに、

ネットワークあるいは実社会上で発生したイベントの有無を検出するためのイベント収集先サイトを登録するイベント収集先サイト登録ステップと、

テキスト、画像、音声等のデータを含むドキュメントを収集するための情報収集先サイトを登録する情報収集先サイト登録ステップと、

前記登録されたイベント収集先サイトから情報を取得し、取得情報の更新の有

無からイベント発生の有無を検出するイベント検出ステップと、

前記イベント検出ステップで検出された更新領域からキーワードを抽出するキーワード抽出ステップと、

前記キーワード抽出部で抽出されたキーワードを用いて前記登録された情報収集先サイト内のドキュメントを検索する情報検索ステップと、

前記情報検索ステップの検索結果をユーザに通知する情報通知ステップと、
を実行させることを特徴とするプログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、インターネット上のサイト情報を自動収集してユーザに通知する情報収集装置、方法及びプログラムに関し、特に、登録したサイトの情報更新を自動的に検出して更新内容に対応したサイト情報を自動収集してユーザに通知する情報収集装置、方法及びプログラムに関する。

【0002】

【従来の技術】

インターネット上には、企業や政府、自治体、個人などさまざまな情報データベース（サイト）が接続されており、インターネットのユーザはこれらの情報データベースから必要な情報、有益な情報を取得することができる。

【0003】

ネットワーク、例えばインターネット上には、テキスト、音声、画像など、各種のデータとそれらを組み合わせた情報（以下「ドキュメント」という）が登録されている。ドキュメントの種類は、広告、案内、マニュアル、ツールなど、非常に多岐に渡っており、特定のユーザにとって必要のないものもあれば、非常に有益なものもある。

【0004】

これらドキュメントの中でも、新しいドキュメントは特に利用価値が高い。例

えば、新しいコンピュータウィルスの発生の通知や、それに対する防御方法、駆除方法などの情報は、インターネットに接続しているユーザにとって貴重な情報である。

【0005】

ネットワークの特徴のひとつは即時性である。ネットワーク上の情報は、タイムラグ無しに取得することができる。コンピュータウィルスに限らず、インターネット上あるいは実社会で発生した事象（以下「イベント」という）の有無をインターネット上のドキュメントから取得することで、ユーザにとって有益な情報をすばやく取得することができる。

【0006】

ネットワーク上のドキュメントを取得する既存システムとしては、例えば検索エンジンがある。検索エンジンは、インターネット上のドキュメントとそのキーワードを共にサーバに登録し、ユーザが入力したキーワードで、情報を検索するシステムであり、エージェント、自動収集ロボットなどと呼ばれているが、インターネット上のサーバに格納されているドキュメントを走査し、表示用ドキュメントと検索用のキーワードデータベースとを生成する。

【0007】

それ以外のネットワーク上のドキュメントを取得する既存システムとしては、情報更新通知システムがある。この情報更新通知システムは、ユーザが指定した特定のページを定期的に監視し、変更があった場合その旨ユーザに通知するシステムである。これには、以下の方法が提案されている。

- (1) 特許第3036445号「ホームページの更新情報監視システム」
- (2) 第3062104号「WWW更新通知システム」
- (3) 特開平10-198614号「ハイパーテキスト文書更新検知方法およびクライアント」
- (4) 特開平11-15716号「文書更新通知装置および文書更新通知方法」
- (5) 特開平11-25020号「WWW掲載番組の内容に変更があったことを依頼者に通知する調査代行サービス装置」
- (6) 特開平11-259354号「インターネット上の情報更新確認方法」

(7) 特開2000-35913 号「ハイパーテキスト文書更新検知方法およびクライアント」

(8) 特開2000-276394号「ウェブページ情報中継システムおよびウェブページ情報中継方法」

(9) 特開2000-357122 号「ウェブページ更新通知方法、記録媒体およびウェブページ更新通知システム」

(10) 特開2001-256100 号「ワールドワイドウェブブラウザ装置およびワールドワイドウェブの更新通知方法」

(11) 特開2002-73455号「Web ページ更新通知方法、クライアントサービスサーバおよびプログラム格納媒体」

これらはいずれも、インターネット上のWWWサイトが更新されたときに、更新されたことをユーザに通知する技術であり、ユーザはキーワードを設定することなく、情報の更新を知ることができる。

【0008】

【発明が解決しようとする課題】

しかしながら、このような従来のネットワーク上のドキュメントを取得するシステムや方法には、それぞれ課題がある。以下、これら従来技術の問題点に説明する。

【0009】

(検索エンジン)

検索エンジンは、予めインターネット上のサイトから情報を取得し、検索用のキーワードを用いてユーザが必要としている情報を抽出する。検索エンジンの第1の問題点は、ユーザがキーワードを設定しなければならないことである。

【0010】

インターネット上の膨大なドキュメントを対象とする検索エンジンでは、特定の情報を取得するためには、正確なキーワードを入力する必要がある。しかし、一般のユーザにとって「自分が欲しい情報」に関連する「キーワード」を適切に設定するのは困難である。

【0011】

例えば、子供の教育に興味のあるユーザが検索エンジンで「子育て」をキーワードに検索した場合、10万件以上のサイトがヒットする。これら検索結果全てにアクセスするのは不可能であるため、通常は他のキーワードを用いて絞込検索をしなければならない。

【0012】

しかし、絞り込み用にキーワードの設定を誤ると、数千件～数万件の検索結果が残り絞り込みできていないことや、逆に絞り込みすぎて目的の情報が検索できなくなるなどの問題が生じる。このように、目的の情報を得るためのキーワードの設定は難しく、一般ユーザには容易に設定できない。

【0013】

検索エンジンの第2の問題点は、ユーザが欲している情報に関する情報を予めユーザが知っておかなければならないという点である。例えば、あるメーカA社が「XXX」という新製品を発売したとする。ユーザがこのA社の「XXX」に関する情報が欲しいと思った場合、ユーザが「A社がXXXを発売した」ということを知っていれば、「XXX」をキーワードに検索エンジンで検索することができる。

【0014】

しかし、ユーザが製品名を知らずに「A社が新製品を発表した」ということしか知らなければ、「XXX」をキーワードにすることはできない。「A社の新製品」で検索すると、「XXX」ではなく、それよりも古い製品のニュースリリース（ニュースリリース発表時は「新製品」）などがヒットする可能性がある。

【0015】

さらに、A社が新製品を出したことすら知らなければ、ユーザはA社の新製品に興味があるにもかかわらず、その情報を取得できない。そのためユーザは定期的にA社のホームページにアクセスして新製品が発表されたかどうかを確認し続ける必要がある。このように、目的の情報を得るためには予め目的の情報に関する事柄をユーザが知っておく必要があり、ユーザが知らない事柄に関する情報を取得することはできない。

【0016】

(WWWページの更新検知)

WWWの更新通知技術は、ユーザが定期的にアクセスする代わりにシステムが情報更新の有無を確認する。以下、既存のWWWページ更新方法の問題点を説明する。

【0017】

(1) 特許第3036445号「ホームページの更新情報監視システム」

このシステムは、WWWのページのチェックサム、ファイルサイズ、ヘッダ情報などをもとに、ドキュメントの更新があったかどうかを判断している。このシステムで分かるのは変更があったということだけであり、どんな変更があったのかはユーザがアクセスして確認する必要がある。

【0018】

(2) 特許第3062104号「WWW更新通知システム」

このシステムでは、WWWサーバのファイルの更新を検出する検出サーバが、ファイルの変更があった際に該ファイルに対応付けられたユーザに対して変更を通知する。このシステムでも前記(1)のシステムと同様、分かるのは変更があったということだけであり、どんな変更があったのかはユーザがアクセスして確認する必要がある。

【0019】

(3) 特開平10-198614号「ハイパーテキスト文書更新検知方法およびクライアント」

この方法では、WWWサーバのファイルの更新をクライアント側がCRCを用いて検出する。この方法でも前記(1)のシステムと同様、分かるのは変更があったということだけであり、どんな変更があったのかはユーザがアクセスして確認する必要がある。

【0020】

(4) 特開平11-15716号「文書更新通知装置および文書更新通知方法」

この装置及び方法では、文書の仲介を行う仲介装置が文書の更新の有無を検出し、検出があった場合、その旨ユーザに通知する。この場合、文書を要求したユ

ーザに対しては、どこが変更されたか強調表示することで変更箇所を分かりやすく表示している。この装置及び方法は、文書の取得要求があったときに更新の有無を判断するため、取得要求の頻度の少ないドキュメントに対しては取得要求が発生するまで文書が更新したかどうか分からない。また、ユーザに通知する内容も、前記(1)～(3)と同様、更新したことだけであり、どんな更新が行われたかはユーザが文書要求して確認するしかない。

【0021】

(5) 特開平11-25020号「WWW掲載番組の内容に変更があったことを依頼者に通知する調査代行サービス装置」

この装置では、WWW番組の内容に変更があったかどうかをユーザに代わって調査する調査代行サーバが、ユーザから要求のあった番組を監視し、変更があった場合そのことを依頼元のユーザに通知するシステムである。この装置でも前記(1)のシステムと同様、分かるのは変更があったということだけであり、どんな変更があったのかはユーザがアクセスして確認する必要がある。

【0022】

(6) 特開平11-259354号「インターネット上の情報更新確認方法」

この方法では、Webサーバ内に文書の更新を監視するWebページ確認サーバを有し、Web確認サーバがサーバレットに登録された情報に基づいてWebページの変更を確認する。この方法でも、前記(1)のシステムと同様、分かるのは変更があったということだけであり、どんな変更があったのかはユーザがアクセスして確認する必要がある。

【0023】

(7) 特開2000-35913号「ハイパーテキスト文書更新検知方法およびクライアント」

この方法では、前記(1)のシステムと同様、ドキュメントのチェックサムを比較してドキュメントの更新の有無を確認する。この方法でも、分かるのは変更があったということだけであり、どんな変更があったのかはユーザがアクセスして確認する必要がある。

【0024】

(8) 特開2000-276394号「ウェブページ情報中継システムおよびウェブページ情報中継方法」

この方法では、ウェブページを中継する中継システムが、ネットワークに対してポーリングを行い情報の更新の有無を確認する。変更があった場合、その変更内容をユーザに通知する。この方法は、前記(1) ～ (7) とは異なり、変更があったことだけでなく変更内容そのものを送信するため、ユーザがアクセスすることなく変更内容の中継システムからの通知により確認することができる。

【 0 0 2 5 】

この方法では、変更内容のみ確認できるが、それ以外の情報、例えば他サーバに格納されている変更内容に関連する情報に関しては、ユーザが他サーバをアクセスする必要がある。

【 0 0 2 6 】

また、インターネット上のドキュメントの場合、変更が頻繁に行われる。例えばニュースサイトなどでは、1 日～2 日程度で変更あるいは削除される場合がある。ユーザが変更通知を受け取っても、実際にアクセスするまでに間がある場合は、すでにドキュメント自体がなくなっている場合がある。

【 0 0 2 7 】

(9) 特開2000-357122号「ウェブページ更新通知方法、記録媒体およびウェブページ更新通知システム」

この方法では、WWWの情報更新を検知するサーバがクライアントに情報更新を伝える際に、電話番号通知機能を使って特定のサーバからの連絡であることを認証している。この方法は、意図していないサーバからの連絡を防ぐことができるため、セキュリティの高いシステムである。

【 0 0 2 8 】

しかし、更新の内容については、前記(1) のシステムと同様、ユーザがアクセスしなければどんな更新がなされたか知ることはできない。

【 0 0 2 9 】

(1 0) 特開2001-256100号「ワールドワイドウェブブラウザ装置およびワールドワイドウェブの更新通知方法」

この方法では、WWWの情報が更新された際に、WWWブラウザにその旨を示す画像を表示することで、ユーザに情報更新を通知する。この方法でも前記（１）のシステムと同様、分かるのは変更があったということだけであり、どんな変更があったのかはユーザがアクセスして確認する必要がある。

【 0 0 3 0 】

（１１）特開2002-73455号「Web ページ更新通知方法、クライアントサービスサーバおよびプログラム格納媒体」

この方法は、予めユーザから更新通知の要求のあったWeb ページの情報とユーザの電子メールアドレスとを保存し、更新を検知したときは該電子メールアドレスにその旨通知するシステムである。この方法でも前記（１）のシステムと同様、分かるのは変更があったということだけであり、どんな変更があったのかはユーザがアクセスして確認する必要がある。

【 0 0 3 1 】

以上のように、従来の方法は、いずれも予め定めたページが更新された際に、そのことをユーザに通知する技術である。即ち（１）～（７）及び（９）～（１１）の従来技術は、更新があったことだけをユーザに通知するだけで、どのような更新が行われたかはユーザが直接アクセスして確認するしかない。

【 0 0 3 2 】

また、（８）の従来技術では、変更内容をユーザに通知するため、ユーザは元情報にアクセスすることなく更新の内容を知ることができる。しかし、この技術でも、知ることができるのは更新があったドキュメント（WWWページ）に関してのみである。

【 0 0 3 3 】

例えば、新製品情報が企業のホームページに登録されたとき、「新製品情報」のページなどを監視することで新製品が登録されたことを知ることはできるが、その新製品の詳細な概要は別の場所に登録されていることが多い。また、その製品の評判が知りたいときは、他のサーバ、例えば技術系ニュースサイトや掲示板サイトなどにアクセスしなければならない。

【 0 0 3 4 】

このように従来技術にあっては、更新された情報のより詳細な情報を得るためには、「更新された」という通知を元に、ユーザが自身で情報を収集しなければならない。

【 0 0 3 5 】

本発明は、ユーザがキーワードなどを設定することなく、未知の情報であっても、ユーザの代わりに複数の情報提供先から情報を収集可能とする情報収集装置、方法及びプログラムを提供することを目的とする。

【 0 0 3 6 】

【課題を解決するための手段】

図 1 は本発明の原理説明図である。本発明は、情報収集装置であって、図 1 (A) のように、ネットワークに接続するネットワーク接続部 1 6 と、ネットワークあるいは実社会上で発生したイベントの有無を検出するためのイベント収集先サイトを登録するイベント収集先サイト登録部 1 8 と、テキスト、画像、音声等のデータを含むドキュメントを収集するための情報収集先サイトを登録する情報収集先サイト登録部 2 0 と、登録されたイベント収集先サイトから情報を取得し、取得情報の更新の有無からイベント発生の有無を検出するイベント検出部 2 2 と、イベント検出部 2 2 で検出された情報の更新領域からキーワードを抽出するキーワード抽出部 2 4 と、キーワード抽出部 2 4 で抽出されたキーワードを用いて登録された情報収集先サイト内のドキュメントを検索する情報検索部 2 6 と、情報検索部 2 6 による検索結果をユーザに通知する情報通知部 2 8 とを備えたことを特徴とする。

【 0 0 3 7 】

このため本発明は、イベント収集先サイトとして特定のサーバ、例えば WWW サイトを監視し、情報の更新によるイベント発生を検出した場合は、その更新内容から新製品発表、新ウィルス発生などのイベントを特定するキーワードを抽出し、このキーワードを使用して情報収集先サイトとして登録したサーバから情報を収集してユーザに自動的に通知する。この結果、ユーザがキーワードなどの情報を特定するための単語を設定することなく、ユーザにとって未知の情報であっ

ても、複数の情報提供先から情報を自動的に収集してユーザに提供できる。

【0038】

ここで、イベント検出部22は、イベント収集先サイトをアクセスしてサイト内のドキュメントをダウンロードしてリファレンスとして保存し、その後、同じイベント収集先サイトからダウンロードしたと共に、このダウンロードドキュメントを用いてリファレンスを更新する。

【0039】

また情報検索部26は、情報収集先サイトをアクセスしてサイト内のドキュメントをダウンロードし、ダウンロードしたドキュメントの中からキーワードを用いて該当するドキュメント部分を検索する。

【0040】

本発明の情報収集装置は、更に、情報検索部26により情報収集先サイトから取得されたドキュメントを保存するドキュメント保存部32を設ける。ドキュメント保存部32は、情報検索部26で検索された検索済みドキュメントを、検索に使用したキーワードをインデックスとして保存する。このため、情報収集先サイトから情報が削除されても、ユーザは必要なドキュメントにいつでもアクセスすることができる。

【0041】

情報検索部26は、情報収集サイト先を定期的にアクセスしてサイト内のドキュメントをダウンロードしてドキュメント保存部32に保存し、その後、イベント検出時にキーワード抽出部で抽出されたキーワードを用いてドキュメント保存部32に保存されたドキュメントを検索する。

【0042】

このため、イベント発生を検出し、関連情報の検索し、ユーザに通知するという順番で処理することを本発明の基本とするが、情報の種類によっては、情報収集先サイトに情報が登録されるのが先で、後日にイベント収集先サイトに情報が登録される場合がある。このような場合には、イベント収集先サイトからイベント発生を検知したときには情報収集先サイトからすでに情報が削除されている場合がある。

【 0 0 4 3 】

そこで、予め情報収集先サイトのドキュメントを外部記憶装置等のドキュメント保存部 3 2 に保存しておき、保存しているドキュメントに対して検索を行うことで、イベント収集先サイトより先に情報収集先サイトに登録された情報であっても、収集することができる。

【 0 0 4 4 】

情報検索部 2 6 は、検索された回数をドキュメントごとに計数し、検索回数が予め定めた閾値以下のドキュメントをドキュメント保存部 3 2 から削除し、新しいドキュメントの保存ができなくなることを防ぐ。ドキュメント削除のタイミングはドキュメント収集時や一定期間ごとに行えば良い。

【 0 0 4 5 】

情報検索部 2 6 は、ドキュメント保存部 3 2 の空き容量の不足を判断した場合、検索回数を判断する閾値を大きくして閾値以下の検索回数のドキュメントをドキュメント保存部 3 2 から削除する。これによって検索回数が予め定めた閾値以下のドキュメントを削除しても外部記憶装置内の空き容量が十分でない場合、閾値を大きくすることで、空き容量を増やすことができる。

【 0 0 4 6 】

イベント検出部 2 2 は、イベント収集先サイトから取得したドキュメントの更新領域に加え、削除された破棄領域を検出し、破棄領域から抽出されたキーワードによりドキュメント保存部 3 2 を検索して保存されたドキュメントから破棄領域を削除する。

【 0 0 4 7 】

このためイベント収集先サイトの情報更新から抽出されたキーワードにより検索されて保存されている情報収集サイトのドキュメントは、古くなってイベント収集先サイトの情報更新により削除されると、削除された破棄領域からキーワードを抽出して保存ドキュメントが自動的に削除され、保存ドキュメントが増えすぎて満杯になることを防ぐ。

【 0 0 4 8 】

情報検索部 2 6 は、キーワード抽出部 2 4 で抽出されたキーワードを用いて一

定期間に亘り定期的に登録された情報収集先サイト内のドキュメントを検索する。これにより次の作用が得られる。イベント収集先サイトからイベント発生を検出して情報収集先サイトからドキュメントの検索を開始した場合、イベント収集先サイトと情報収集先サイトが異なる場合は、各サイトへの情報登録のタイミングが異なる場合がある。

【 0 0 4 9 】

この場合、イベントを検出して情報収集を開始しても情報収集先サイトにはまだ登録されておらず、必要な情報が得られない。そこで、一定期間に亘り定期的に情報検索を繰り返すことで、イベント収集先サイトと情報収集先サイトでの登録のタイミングのずれによる情報収集漏れを防ぐ。

【 0 0 5 0 】

情報検索部 2 6 は、キーワードを用いたドキュメントの検索回数を計数し、一定期間経過した時のドキュメントの検索回数が予め定めた閾値を超えた場合は、キーワードによるドキュメントの情報検索を再度一定期間継続し、閾値以下であった場合は、キーワードによる情報検索を停止する。これにより次の作用が得られる。

【 0 0 5 1 】

イベント収集先サイトと情報収集先サイトでの登録のタイミングのずれがあると、定期的に検索してもタイムラグの長さによってはやはり情報を取得できない場合がある。そこで、検索回数を保存し、一定期間の検索回数が予め定めた閾値以下であれば、イベントの新規性が薄れてきたと判断して情報収集を停止する。

【 0 0 5 2 】

イベント収集先サイト登録部 1 8 は、ネットワークを介してイベント収集先リストサーバからイベント収集先サイトを取得して登録し、情報収集先サイト登録部 2 0 は、ネットワークを介して情報収集先リストサーバから情報収集先サイトを取得して登録する。本発明は、イベント収集先サイトや情報収集先サイトを予め登録するが、専用のサーバからリストを取得して登録するようにしても良い。

【 0 0 5 3 】

またイベント収集先サイト登録部 1 8 は、ネットワークを介して同一構成を備

えた他の情報収集装置からイベント収集先サイトを取得して登録し、同様に、情報収集先サイト登録部 2 0 は、ネットワークを介して同一構成を備えた情報収集装置から情報収集先サイトを取得して登録するようにしても良い。これは本発明の情報収集装置は、インターネットを介して接続されているコンピュータ上に存在していることから、同じ情報収集装置の間で相互にイベント収集先サイトや情報収集先サイトを利用し合う形態、所謂ピア・ツー・ピアとして利用する。

【 0 0 5 4 】

キーワード抽出部 2 4 は、イベント検出部で検出された情報の更新領域を形態素解析して品詞ごとに分割した後に固有名詞のみを抽出し、抽出した名詞がキーワードデータベース 3 0 に登録されている既存のキーワードと異なる場合に、情報検索部 2 6 に抽出した固有名詞をキーワードとして出力する。これによってイベント収集先サイトの更新情報から例えば新製品名や新しいコンピュータウィルス名などをキーワードとして出力し、このキーワードによる情報収集先サイトからのドキュメント検索による情報収集ができる。

【 0 0 5 5 】

キーワード抽出部 2 4 は、情報検索部 2 6 にキーワードとして出力した固有名詞をキーワードデータベース 3 0 に追加登録する。これにより今回のイベントで抽出したキーワードをキーワードデータベース 3 0 に追加登録して、次回以降のイベントで再度キーワードとして抽出されることを防ぎ、検索の済んだキーワードによる検索が不必要に行われることを回避する。

【 0 0 5 6 】

キーワード抽出部 2 4 は、イベント検出部 2 2 で検出された情報の更新領域から複数のキーワードを抽出した場合、更新領域の内容に基づいて各キーワードの優先順位を付与して情報検索部 2 6 に出力する。

【 0 0 5 7 】

イベント検出部 2 2 は、イベント発生を検出したイベント収集先サイトの更新領域に新規情報のみの追加があった場合は、新規情報の履歴を保存し、更新領域に新規情報が追加されると同時に古い情報が削除された場合は、新規情報の履歴と削除情報の履歴を保存し、情報通知部 2 8 により保存履歴をユーザに通知可能

とする。

【 0 0 5 8 】

この更新履歴の保存によって、イベント収集先サイトの更新された情報の一覧等をユーザに通知でき、どのような時系列で情報が更新、削除されたかを知ることができる。例えば新規情報と削除情報をマージすることによって、例えば過去から現在まで開発した製品の一覧と、現在まだ扱っている製品一覧を得ることができる。

【 0 0 5 9 】

イベント検出部 2 2 は、イベント発生を検出したイベント収集先サイトの更新領域に新規情報のみの追加が合った場合は、新規情報の履歴としてキーワード抽出部 2 4 で抽出されたキーワードを保存し、また更新領域に新規情報が追加されると同時に古い情報が削除された場合は、新規情報の履歴と削除情報の履歴としてキーワード抽出部 2 4 で抽出されたキーワードを保存し、情報通知部 2 8 により保存履歴したキーワードをユーザに通知可能とする。

【 0 0 6 0 】

このためキーワードを抽出して更新履歴としてキーワードの一覧をユーザに通知することで、更新領域の履歴のみを並べた場合に比べ、履歴を容易に把握できる。

【 0 0 6 1 】

イベント検出部 2 2 は、更新領域に追加された新規情報に外部サイトのリンクがあった場合、外部リンク先のドキュメントをダウンロードしてドキュメント保存部 3 2 に保存すると共に、新規情報の履歴にドキュメント保存部 3 2 の保存ドキュメントをリンクさせる。これにより情報収集先サーバからドキュメントが削除されても、ユーザは常にドキュメントにアクセスすることができる。

【 0 0 6 2 】

本発明は、ネットワーク環境を対象とした情報収集方法を提供する。この情報収集方法は、

イベント収集先サイト登録部により、ネットワークあるいは実社会上で発生したイベントの有無を検出するためのイベント収集先サイトを登録するイベント収

集先サイト登録ステップと、

情報収集先サイト登録部により、テキスト、画像、音声等のデータを含むドキュメントを収集するための情報収集先サイトを登録する情報収集先サイト登録ステップと、

イベント検出部により、登録されたイベント収集先サイトから情報を取得し、取得情報の更新の有無からイベント発生の有無を検出するイベント検出ステップと、

キーワード抽出部により、イベント検出ステップで検出された情報の更新領域からキーワードを抽出するキーワード抽出ステップと、

情報検索部により、キーワード抽出ステップで抽出されたキーワードを用いて登録された情報収集先サイト内のドキュメントを検索する情報検索ステップと、

情報通知部により、情報検索ステップの検索結果をユーザに通知する情報通知ステップと、

を備えたことを特徴とする。

【 0 0 6 3 】

また本発明は、コンピュータで実行されるプログラムを提供する。このプログラムは、コンピュータに、

ネットワークあるいは実社会上で発生したイベントの有無を検出するためのイベント収集先サイトを登録するイベント収集先サイト登録ステップと、

テキスト、画像、音声等のデータを含むドキュメントを収集するための情報収集先サイトを登録する情報収集先サイト登録ステップと、

登録されたイベント収集先サイトから情報を取得し、取得情報の更新の有無から判別してイベント発生の有無を検出するイベント検出ステップと、

前記イベント検出ステップで検出された情報の更新領域からキーワードを抽出するキーワード抽出ステップと、

キーワード抽出ステップで抽出されたキーワードを用いて登録された情報収集先サイト内のドキュメントを検索する情報検索ステップと、

情報検索ステップの検索結果をユーザに通知する情報通知ステップと、
を実行させることを特徴とする。

【0064】

本発明による情報収集方法及びプログラムの詳細は、情報収集装置の場合と基本的に同じになる。

【0065】

【発明の実施の形態】

図2は、本発明による情報収集装置の実施形態を、本発明が適用されるネットワーク環境と共に示した機能ブロック図である。

【0066】

図2において、本発明の情報収集装置10は、例えばユーザが保有しているパーソナルコンピュータにより実現されており、インターネット11などのネットワークに接続されてインターネット上に構築されている情報データベースとして機能するサイトからユーザに必要な情報を収集して利用する。

【0067】

本発明の情報収集装置10にあっては、インターネット11に接続されている各種のサーバ、例えばftpサーバ、WAISサーバ、Archieサーバ、WWWサーバ、NEWSサーバをアクセス対象とすることができ、この実施形態にあってはWWWサーバを例にとって説明する。

【0068】

本発明にあっては、インターネット上あるいは実社会で発生した事象をイベントと定義しており、このイベントの有無をインターネット上のサイトから取得することで、ユーザにとって有益な情報を収集する。このため本発明にあっては、イベント発生の有無を検出する対象となるサーバをイベント収集先サイトと呼んでおり、図2の例にあってはインターネット11に接続されているWWWサーバで構築されているイベント収集先サイト12-1、12-2、12-3をイベント発生を検出先としている。

【0069】

また本発明にあっては、特定の情報を収集するWWWサーバを情報収集先サイトと定義しており、図2の例にあってはWWWサーバで実現される3つの情報収

集先サイト14-1, 14-2, 14-3を例にとって示している。このイベント収集先サイト12-1~12-3と情報収集先サイト14-1~14-3は、別のWWWサーバであってもよいし、同じWWWサーバであってもよい。

【0070】

本発明の情報収集装置10は、ネットワーク接続部16、イベント収集先サイト登録部18、情報収集先サイト登録部20、イベント検出部22、キーワード抽出部24、情報検索部26、情報通知部28、キーワードデータベース30、ドキュメント保存部32及び表示部34を備えている。

【0071】

イベント収集先サイト登録部18は、イベント発生の有無を検出するためのイベント収集先サイト12-1~12-3を登録している。具体的には、イベント収集先サイト12-1~12-3のアドレスとなるURLを登録している。イベント収集先サイトとしては、ユーザが情報収集を必要とする任意のサイトを検索もしくは収集して、予め登録している。

【0072】

情報収集先サイト登録部20は、テキスト、画像、音声などのデータを含む情報を収集するための情報収集先サイト14-1~14-3を予め登録する。ここで、本発明の情報収集装置10が収集するインターネット上のテキスト、画像、音声などを含む情報をドキュメントと定義する。情報収集先サイト登録部20にもイベント収集先サイト登録部18と同様、情報収集先サイト14-1~14-3のURLを例えばユーザが予め調べて登録している。

【0073】

イベント検出部22は、イベント収集先サイト登録部18に登録されたイベント収集先サイト12-1~12-3から情報を取得し、取得した情報の変化領域となる更新の有無からイベント発生の有無を検出し、イベント収集先サイトの情報に変化があったことを、情報通知部28を介して表示部34に表示してユーザに通知する。

【0074】

キーワード抽出部24は、イベント検出部22で検出されたイベント収集先サ

イトの情報の更新領域、即ち変化のあった領域からキーワードを抽出する。このキーワード抽出は更新領域のテキスト文書の形態素解析により、例えば名詞としてのキーワードを抽出する。

【 0 0 7 5 】

キーワード抽出部 2 4 に対し設けているキーワードデータベース 3 0 は、過去のイベント検出で抽出された使用済みのキーワードが登録されている。このためキーワード抽出部 2 4 は、新たなイベント検出でキーワードを抽出した際にキーワードデータベース 3 0 を参照し、もし既に登録されたキーワードと同じであれば、抽出したキーワードによる情報収集は済んでいることからキーワードを破棄し、一方、キーワードデータベース 3 0 に登録されていないならば、新たなキーワードとして情報検索部 2 6 に出力する。

【 0 0 7 6 】

情報検索部 2 6 は、キーワード抽出部 2 4 で検出されたキーワードを用いて、情報収集先サイト登録部 2 0 に登録されている情報収集先サイト 1 4 - 1 ~ 1 4 - 3 のドキュメントを検索し、キーワードを含むドキュメントを取得する。

【 0 0 7 7 】

更に情報通知部 2 8 は、情報検索部 2 6 によりキーワードに基づいて検索された検索結果としての情報収集先サイト 1 4 - 1 ~ 1 4 - 3 のいずれかから得られた検索結果としてのドキュメントの存在を表示部 3 4 に表示してユーザに通知する。

【 0 0 7 8 】

また情報収集装置 1 0 の情報検索部 2 6 に対してはドキュメント保存部 3 2 が設けられている。ドキュメント保存部 3 2 は、情報検索部 2 6 の検索結果として得られたドキュメント、あるいは情報収集処理に先立ってイベント収集先サイト登録部 1 8 の登録先となるイベント収集先サイト 1 2 - 1 ~ 1 2 - 3 から予め取得したドキュメントなどを保存する。

【 0 0 7 9 】

このドキュメント保存部 3 2 は、保存先としてハードディスクドライブ HDD を使用しており、併せてハードディスクドライブ HDD に対する保存制御の機能

を備えている。この点はイベント収集先サイト登録部 1 8、情報収集先サイト登録部 2 0、更にキーワードデータベース 3 0 についても同様であり、それぞれ保存先としてハードディスクドライブ HDD の領域を割り当てており、これに加えて登録制御、データベースアクセスの制御機能を併せて備えている。

【 0 0 8 0 】

更に、図 2 のインターネット 1 1 に対しては本発明の情報収集装置 1 0 と同じ構成を備えた情報収集装置 1 0 - 1、1 0 - 2 が接続されており、これは他のユーザが使用している本発明の情報収集装置である。

【 0 0 8 1 】

またインターネット 1 1 に対しては情報収集先リストサーバ 1 5 - 1 やイベント収集先リストサーバ 1 5 - 2 が接続されている場合があり、本発明の情報収集装置 1 0 にあっては情報収集先サイト及びイベント収集先サイトの登録の際に、情報収集先リストサーバ 1 5 - 1 やイベント収集先リストサーバ 1 5 - 2 にアクセスして、それぞれの収集先のリストを収集して、情報収集先サイト登録部 2 0 やイベント収集先サイト登録部 1 8 に対する登録を行うことができるようにしている。

【 0 0 8 2 】

図 2 における本発明の情報収集装置 1 0 は、例えば図 3 のようなコンピュータのハードウェア資源により実現される。

【 0 0 8 3 】

図 3 のコンピュータにおいて、CPU 1 0 0 のバス 1 0 1 には RAM 1 0 2、ハードディスクコントローラ (ソフト) 1 0 4、フロッピーディスクドライバ (ソフト) 1 1 0、CD-ROM ドライバ (ソフト) 1 1 4、マウスコントローラ 1 1 8、キーボードコントローラ 1 2 2、ディスプレイコントローラ 1 2 6、通信用ボード 1 3 0 が接続される。

【 0 0 8 4 】

ハードディスクコントローラ 1 0 4 はハードディスクドライブ 1 0 6 を接続し、本発明の情報収集処理を実行するアプリケーションプログラムをローディングしており、コンピュータの起動時にハードディスクドライブ 1 0 6 から必要なプ

プログラムを呼び出して、RAM102上に展開し、CPU100により実行する。

【0085】

フロッピーディスクドライバ110にはフロッピーディスクドライブ（ハード）112が接続され、フロッピーディスク（R）に対する読み書きができる。CD-ROMドライバ114に対しては、CDドライブ（ハード）116が接続され、CDに記憶されたデータやプログラムを読み込むことができる。

【0086】

マウスコントローラ118はマウス120の入力操作をCPU100に伝える。キーボードコントローラ122はキーボード124の入力操作をCPU100に伝える。ディスプレイコントローラ126は表示部34に対して表示を行う。通信用ボード130は無線を含む通信回線132を使用し、インターネット等のネットワークを介して他のコンピュータやサーバとの間で通信を行う。

【0087】

図4は、図2の情報収集装置10による本発明の情報収集処理の基本的な処理手順を示したフローチャートであり、このフローチャートが本発明による情報収集のためのアプリケーションプログラムの実施形態に対応する。

【0088】

図4において、まずステップS1でイベント収集先サイト登録部18にイベント収集先サイトを登録する。ここではイベント収集先サイトとして、例えばA社のトピックスのページのURLを登録する。このA社のトピックスのURLを使用してイベント収集先サイトにアクセスすると、例えば図5（A）のような新製品情報に関するドキュメント36-1が取得できる。

【0089】

続いてステップS2で、情報収集先サイト登録部20に情報収集先サイトを登録する。この情報収集先サイトはA社のホームページであってもよいし、A社と同じ業種を含む製品紹介などを行っている他の情報収集先サイトなどを登録してもよい。

【0090】

次にステップ S 3 で、イベント収集先サイトである A 社のトピックスのページにアクセスし、図 5 (A) のような新製品情報のドキュメント 3 6 - 1 をダウンロードし、これをリファレンスとして保存する。リファレンスとして保存される図 5 (A) の新製品情報のドキュメント 3 6 - 1 は、例えば新製品「AAA」～「FFF」について、発売開始とその年月日が記述されている。

【0091】

続いてステップ S 4 で登録したイベント収集先サイトに定期的にアクセスしてドキュメントをダウンロードし、ステップ S 5 で保存ページであるリファレンスと取得ページを比較し、ステップ S 6 で変化があるかどうか判断する。

【0092】

この定期的なイベント収集先サイトのページのダウンロードで、例えば図 5 (B) のような新製品情報のドキュメント 3 6 - 2 が取得されたとする。この新製品情報のドキュメント 3 6 - 2 は、図 5 (A) のリファレンスとしてのドキュメント 3 6 - 1 と対比すると、リファレンスのドキュメント 3 6 - 1 の一番下の最も古い新製品「AAA」に関する情報 3 8 が削除されて、先頭に新製品「XXX」の情報 4 0 が追加されている。

【0093】

ここで図 5 (A) のリファレンスとしてのドキュメント 3 6 - 1 から削除されている最も古い情報 3 8 を破棄領域とする。また、図 5 (B) のドキュメント 3 6 - 2 において新たに追加された新規な情報 4 0 を更新領域という。

【0094】

このようにして図 5 (A) のリファレンスとしてのドキュメント 3 6 - 1 に対し、新たに取得したリファレンス 3 6 - 2 に変化があった場合、ステップ S 7 で、取得した図 5 (B) のドキュメント 3 6 - 2 の更新領域となる新規な情報 4 0 を抽出し、イベント発生をユーザに通知した後、ステップ S 8 で保存ページとしてのリファレンスを更新する。

【0095】

次にステップ S 9 で、図 5 (B) の更新領域の新規な情報 4 0 を対象に、キーワード抽出部 2 4 が、検出したイベント発生を特定するキーワードを抽出する。

この例にあっては、キーワードとして新製品名である「XXX」がキーワードとして抽出される。

【0096】

このようにして抽出されたキーワードは情報検索部26に与えられ、次のステップS10で情報検索部26が抽出されたキーワードにより登録されている情報収集先サイトのドキュメントを検索し、ステップS11で検索結果を情報通知部28により表示部34に表示してユーザに通知する。

【0097】

このキーワードによる情報検索としては、イベント発生により抽出されたキーワードであるA社の製品名「XXX」を使用した検索によって、A社のサイトにはない「XXX」の評判、レビュー、欠点、小売価格などの情報を自動的に収集してユーザに提供することができる。

【0098】

また本発明の情報収集装置10を使用してユーザがコンピュータウイルスに関する情報収集をしたい場合には、ステップS1においてイベント収集先サイトにアンチウイルスソフト開発会社のURLを予め登録し、またステップS2の情報収集先サイトにパソコンメーカーのホームページを登録しておく。

【0099】

これによりイベント収集先サイトのアクセスによるイベント発生の検出で新しいウイルスの発生を検出し、この新しいウイルスの発生検出により抽出したキーワード例えばウイルス名などによる情報収集先サイトの検索で、新しいウイルスに対しパソコンユーザとしてどのように対処すべきかといった有益な情報を自動的に収集して、ユーザに示すことができる。

【0100】

このように本発明の情報収集装置にあっては、イベント収集先として特定のサイトを監視し、このイベント収集先サイトの情報に更新があった場合は、その更新内容から新製品発表や新ウイルス発生などのイベントを特定するキーワードを生成し、このキーワードにより情報収集先サイトからキーワードを含む情報を収集することにより、ユーザはキーワードなどの情報を特定するための単語を設定

する必要がなく、このためユーザにとって未知の情報であっても、ユーザの代わりに情報収集装置 1 0 が複数の情報提供先から必要な情報を収集してユーザに提供することができる。

【 0 1 0 1 】

ここで、イベント収集先サイトにおける新規情報の追加による更新の形態としては、図 5 (A) のように、最も古い情報 3 8 を削除し図 5 (B) のように新規な情報 4 0 を追加する形態の他に、図 6 のように古い情報を削除することなく新規情報を追加する場合もある。

【 0 1 0 2 】

図 6 (A) は、図 5 (A) と同じ最初に得られた新製品のドキュメント 3 6 - 1 1 であり、これに続いて新製品の情報 4 0 の追加により図 6 (B) のようなドキュメント 3 6 - 1 2 が得られる。この追加した情報 4 0 を含むドキュメント 3 6 - 1 2 は、最も古い新製品「 A A A 」の情報 3 8 を削除せずに残し、先頭に新製品「 X X X 」の情報 4 0 を追加している。もちろんサイトによっては、図 5 と図 6 の両者を組み合わせた新規情報の更新の形態をとる場合もある。

【 0 1 0 3 】

図 7 は、図 2 の情報収集装置 1 0 の他の実施形態における処理手順のフローチャートである。この図 7 の実施形態にあっては、情報検索部 2 6 でイベント発生検出に基づいてキーワード抽出部 2 4 で抽出されたキーワードを用いた情報収集先サイトの検索で得られたドキュメントを、ドキュメント保存部 3 2 に保存するようにしたことを特徴とする。

【 0 1 0 4 】

即ち、図 7 のステップ S 1 ~ S 1 0 は図 4 と同じであるが、ステップ S 1 1 で情報検索部 2 6 によりキーワードを用いて得られたドキュメントをドキュメント保存部 3 2 に保存している。この検索により収集されたドキュメントの保存の際には、検索に使用したキーワードと収集したドキュメントをリンクさせてドキュメント保存部 3 2 に保存する。

【 0 1 0 5 】

このようにキーワードに基づいて検索されたドキュメントを情報収集先サイト

からダウンロードしてハードディスクドライブなどの外部記憶装置で構成されるドキュメント保存部 3 2 に保存することにより、情報収集先サイトから情報が、その後に削除されても、ユーザは必要なドキュメントについて、情報収集装置 1 0 自身のドキュメント保存部 3 2 に対し例えばキーワードをインデックスに使用したアクセスで、いつでも利用できる。

【 0 1 0 6 】

図 8 は、図 2 の情報収集装置 1 0 における他の実施形態の処理手順のフローチャートである。この図 8 の実施形態にあつては、イベント検出による情報検索に先立って、まず情報収集先サイト 1 4 - 1 ~ 1 4 - 3 からドキュメントを取得してドキュメント保存部 3 2 に保存しておき、イベント検出部 2 2 でイベント発生を検出した場合、キーワード抽出部 2 4 で抽出したキーワードを用いて、情報検索部 2 6 がドキュメント保存部 3 2 に保存しているドキュメントを対象に情報検索を行うようにしたことを特徴とする。

【 0 1 0 7 】

図 8 の情報収集処理にあつては、ステップ S 1 でイベント収集先サイトを登録した後、ステップ S 2 で情報収集先サイトを登録すると、ステップ S 3 で登録した情報収集先サイトからドキュメントを取得してドキュメント保存部 3 2 に保存する。

【 0 1 0 8 】

これによってステップ S 3 以降にあつては、ネットワーク上の情報収集先サイトから改めてドキュメントを取得することなく、情報収集装置 1 0 自身のドキュメント保存部 3 2 に保存している情報収集先のドキュメントを対象に、イベント発生に基づく情報検索を行うことになる。

【 0 1 0 9 】

即ちステップ S 4 ~ S 1 2 の処理により、図 4 のステップ S 3 ~ S 1 1 の場合と同様に、イベント発生の検出、イベント発生の検出による変化領域の抽出、変化領域からのキーワードの抽出、キーワードを用いたドキュメント保存部 3 2 の保存ドキュメントを対象とした検索、及び検索結果のユーザ通知を行うことになる。

【0110】

このような図8の情報収集先サイトのドキュメントを予め保存して検索する処理は、情報の種類によっては情報収集先サイトに情報を登録するのが先で後日にイベント収集先サイトに情報が登録される場合に好適である。

【0111】

イベント収集先サイトでイベント発生を検出したときには先に情報が登録された情報収集先サイトから既に対応する情報が削除されている場合、図8の実施形態にあつては、予め情報収集先サイトの情報をドキュメント保存部32に保存した後にイベント発生を検出してドキュメント保存部32内の保存ドキュメントを対象に検索を行っているため、ネットワーク上の情報収集先サイトで既に情報が削除された後であっても、確実にイベント発生に基づいたキーワードによる情報検索を行ってユーザに提供することができる。

【0112】

図9及び図10は、図8の実施形態のように、情報収集先サイトのドキュメントを予めドキュメント保存部32に保存した後にイベント発生に基づく情報収集を行う実施形態について、ドキュメント収集を継続するとドキュメント保存部32を構築しているハードディスクドライブなどの外部記憶装置が満杯となり、新しいドキュメントが保存できなくなることから、これを回避するために定期的にドキュメントを削除する処理を加えている。

【0113】

図9において、ステップS1～S11は図8の実施形態と同じであるが、図9に続く図10のステップS12～S14の処理において、ドキュメント保存部32からドキュメントを削除する処理を行う。

【0114】

即ちステップS12で、情報検索部26により検索したドキュメントの検索回数を計数しており、ステップS3で検索回数が閾値以下のドキュメントをドキュメント保存部32から削除する。例えばステップS13における閾値を0とし、検索回数が0のドキュメントをドキュメント保存部32から削除する。

【0115】

このステップ S 1 2, S 1 3 における検索回数の計数と削除のタイミングは、別のタイミングで行ってもよい。またステップ S 1 3 の削除はドキュメントの収集時に行ってもよいし、別にタイマを保持して一定時間ごとに行ってもよい。

【 0 1 1 6 】

図 1 1 及び図 1 2 は、保存ドキュメントを削除する他の実施形態を含む本発明の情報収集処理のフローチャートである。この実施形態にあっては、検索回数が予め定めた閾値以下のドキュメントを削除してもドキュメント保存部 3 2 の空き容量が十分でない場合に、閾値の値を大きくすることでドキュメント保存部 3 2 における空き容量を増やすようにしたことを特徴とする。

【 0 1 1 7 】

図 1 1 において、ステップ S 1 ~ S 1 1 は図 8 の実施形態と同じであるが、図 1 2 のステップ S 1 2 ~ S 1 7 による処理で空き容量を増やすように検索回数の閾値を変更する処理を行う。

【 0 1 1 8 】

即ちステップ S 1 2 で検索したドキュメントの検索回数を計数した後、ステップ S 1 3 でドキュメント保存部 3 2 の空き容量は十分か否かチェックする。空き容量が不足している場合にはステップ S 1 4 に進み、閾値を例えば 1 つ大きくする。

【 0 1 1 9 】

この閾値は初期状態にあっては例えば 0 であることから、ステップ S 1 4 では閾値は 1 となる。続いてステップ S 1 5 で検索回数が大きくした閾値以下のドキュメントを削除する。これによって閾値 0 で削除していたドキュメントの数に対し、閾値を 1 つ増加したことで、削除するドキュメントの数を増やすことができ、ドキュメントの削除による空き容量を大きくすることができる。

【 0 1 2 0 】

ステップ S 1 5 でドキュメントの削除を行うと、ステップ S 1 6 で、このときの検索結果をユーザに通知した後、再びステップ S 1 3 に戻り、空き容量は十分か否かチェックする。もちろん、空き容量が十分か否かは予め定めた空き容量の閾値を使用して判断する。

【 0 1 2 1 】

空き容量が十分でなければ、ステップ S 1 4 ～ S 1 6 の処理を繰り返す。空き容量が十分に確保できた場合には、ステップ S 1 7 で再び閾値を初期値である 0 に戻した後、図 1 1 のステップ S 3 からの処理を繰り返す。

【 0 1 2 2 】

図 1 3 及び図 1 4 は、ドキュメント保存部からドキュメントを削除する本発明の情報収集装置における他の処理手順の実施形態のフローチャートである。この実施形態にあつては、図 5 (A) (B) のようなイベント収集先サイトから得られたリファレンスとしてのドキュメント 3 6 - 1 と新規情報を含むドキュメント 3 6 - 2 の比較で判断される破棄領域として削除された情報 3 8 に対応した保存ドキュメントを削除するようにしたことを特徴とする。

【 0 1 2 3 】

図 1 3 のステップ S 1 ～ S 1 1 は図 8 のステップ S 1 ～ S 1 1 と同じであり、これに続いて図 1 4 のステップ S 1 2 ～ S 1 4 で図 5 (A) の削除した情報 3 8 に対応するドキュメントの削除処理を行っている。

【 0 1 2 4 】

即ち、ステップ S 1 2 でイベント収集先のページ更新により削除された情報、例えば図 5 (A) の情報 3 8 から、キーワードとして「AAA」を抽出する。続いてステップ S 1 3 で、抽出したキーワード「AAA」を用いて、ドキュメント保存部 3 2 に保存している情報収集先サイトのドキュメントを検索する。これによって、保存されているキーワード「AAA」に対応したドキュメントが検索され、ステップ S 1 3 でドキュメント保存部 3 2 から削除する。

【 0 1 2 5 】

このような図 1 3、図 1 4 の保存ドキュメントの削除処理により、ドキュメント保存部 3 2 に保存されているドキュメントの中から、イベント発生の検出によりイベント収集先サイトから削除された情報に対応する古いドキュメントを自動的に削除していくことができる。

【 0 1 2 6 】

図 1 5 は、図 2 の情報収集装置 1 0 における本発明の情報収集処理の他の実施

形態の処理手順のフローチャートである。この実施形態にあつては、イベント発生を検出により抽出したキーワードを用いた情報収集先サイトに対する情報検索を、一定期間継続して定期的に行うようにしたことを特徴とする。

【 0 1 2 7 】

図 1 5 において、ステップ S 1 ～ S 1 1 の処理は図 4 のステップ S 1 ～ S 1 1 と同じであるが、これに加えてステップ S 1 2 で予め定めた一定期間が経過したか否かチェックしている。一定期間を経過するまではステップ S 1 0, S 1 1 のキーワードによる情報収集先のドキュメントの検索を定期的に繰り返してユーザーに通知するようにしている。

【 0 1 2 8 】

この図 1 5 の処理は、イベント収集先サイトと情報収集先サイトが異なる場合の各サイトにおける情報登録のタイミングずれに対応する。即ち、イベント収集先サイトからイベント発生を検出しても、情報収集先サイトにまだ情報が登録されておらず、必要な情報が得られない場合がある。

【 0 1 2 9 】

このような場合、図 1 5 の実施形態にあつては、ステップ S 1 2 で一定期間経過するか否か判断することで、ステップ S 1 0, S 1 1 の処理の繰り返しによりキーワードを用いた情報検索を繰り返すことで、情報収集先サイトに対する情報登録のタイミングずれによる情報収集漏れを防ぐことができる。

【 0 1 3 0 】

図 1 6 及び図 1 7 は、図 1 5 の実施形態ではカバーできない情報収集先サイトに対する情報登録のタイミングずれによる情報収集漏れを防ぐようにした本発明の情報収集処理の他の実施形態のフローチャートである。

【 0 1 3 1 】

即ち、図 1 5 の実施形態にあつては、キーワードによる情報検索を一定時間経過するまで定期的に繰り返すことにより、情報収集先サイトの情報登録によるタイムラグがあっても情報収集漏れを防ぐようにしているが、タイムラグの長さによつては、やはり情報を収集できない場合がある。

【 0 1 3 2 】

そこで図16及び図17の実施形態にあつては、キーワードを用いた情報検索結果の回数を保持し、一定期間の検索回数が予め定めた閾値以下であれば、イベントの新規性が薄れてきたと判断して、キーワードを用いた情報収集を停止するようにしている。

【0133】

図16のステップS1～S11は図15のステップS1～S11と同じであるが、これに続く図17のステップS12～S14の処理でイベントの新規性が薄れてきたことを判断して、情報収集を停止する。即ち、ステップS12で検索回数の履歴を計数して保存し、ステップS13で一定期間経過した否かチェックし、一定期間経過した場合にはステップS14で検索回数が閾値以下か否かチェックする。

【0134】

検索回数が閾値を超えていればイベントの新規性が高いものと判断し、図16のステップS10からのキーワードによる情報収集先サイトのドキュメントの検索を繰り返す。

【0135】

ステップS14で検索回数が閾値以下となった場合には、イベントの新規性が薄れてきたと判断し、ステップS10からのキーワードによる情報収集先サイトのドキュメント検索を停止して、図16のステップS4に戻って新たなイベント収集先サイトの情報変換の検索処理から繰り返す。

【0136】

尚、図17のステップS13の一定期間の経過を判断する処理を除き、ステップS12で検索結果の履歴を計数して保存し、この検索回数が閾値以下となった場合には、直ちに情報検索を停止して、図16のステップS4に戻るようにしても良い。

【0137】

図18及び図19は、図2の情報処理装置10における本発明による情報収集処理の他の実施形態のフローチャートであり、この実施形態にあつてはインターネット上のサーバからイベント収集先サイトや情報収集先サイトの情報を取得す

るようにした事の特徴とする。

【 0 1 3 8 】

図 1 8 及び図 1 9 の実施形態にあっては、図 2 のインターネット 1 1 に接続しているイベント収集先リストサーバ 1 5 - 2 や情報収集先リストサーバ 1 5 - 1 を利用する。インターネットにあっては WWW サーバのアドレス (URL) の変更や、サーバそのものの廃止などが頻繁に起こりうる。

【 0 1 3 9 】

そこでイベント収集先リストサーバ 1 5 - 2 において、イベント収集先サイトを設定し、その情報をクライアントである本発明の情報収集装置 1 0 に提供することで、クライアントとしての情報収集装置 1 0 のユーザーはイベント収集先サイトがどのサーバにあるかなどを一切気にすることなく、イベント収集先リストをイベント収集先サイト登録部 1 8 に登録することができる。

【 0 1 4 0 】

この点は情報収集先サイト登録部 2 0 のサイト登録についても同様であり、情報収集先リストサーバ 1 5 - 1 で情報収集先サイトを設定し、その情報をクライアントとしての情報収集装置 1 0 に提供することで、情報収集先サイトのサーバの状態を意識することなく情報収集先サイトを情報収集サイト登録部 2 0 に登録して情報検索を利用することができる。

【 0 1 4 1 】

このようなイベント収集先リストサーバ 1 5 - 2 及び情報収集先リストサーバ 1 5 - 1 に対応し、図 1 8 の処理にあっては、まずステップ S 1 で情報収集先リストサーバ 1 5 - 1 から情報収集先サイトの情報を取得し、ステップ S 2 で情報収集先サイト登録部 2 0 の登録内容と比較して変更があれば、ステップ S 3 で情報収集先サイト登録部 2 0 に新たな情報収集先サイトの URL を登録する。

【 0 1 4 2 】

またステップ S 4 でイベント収集先リストサーバ 1 5 - 2 からのイベント収集先サイトの情報を収集し、イベント収集先サイト登録部 1 8 の登録内容と比較して、イベント収集先サイトに変更があれば変更のあった新たなイベント収集先サイトをステップ S 6 でイベント収集先サイト登録部 1 8 に登録し、更にステップ

S 7 で新たに登録したイベント収集先サイトのページをリファレンスとして保存する。

【 0 1 4 3 】

これに続くステップ S 8 ～ S 1 5 の処理は、図 4 のステップ S 4 ～ S 1 1 の処理と同じである。

【 0 1 4 4 】

尚、図 1 8 及び図 1 9 の実施形態にあつては、情報収集先リストサーバ 1 5 - 1 とイベント収集先リストサーバ 1 5 - 2 の両方からサイトの情報を取得して登録するようにしているが、いずれか一方についてサーバから情報を取得してサイト登録を行うようにしても良い。

【 0 1 4 5 】

図 2 0 及び図 2 1 は、図 2 の情報収集装置 1 0 における本発明の情報収集処理の他の実施形態のフローチャートであり、この実施形態にあつては図 2 でインターネット 1 1 に接続している本発明と同じ構成を持つ他の情報収集装置 1 0 - 1 , 1 0 - 2 からイベント収集先サイト及び情報収集先サイトの情報を取得するようにしたことを特徴とする。

【 0 1 4 6 】

この図 2 0 及び図 2 1 の実施形態において、本発明の情報収集装置 1 0 が同じ構成を持つ他の情報収集装置 1 0 - 1 , 1 0 - 2 からイベント収集先サイト及び情報収集先サイトの情報を収集するネットワーク環境は、情報収集装置 1 0 , 1 0 - 1 , 1 0 - 2 がそれぞれピアマシンとして相互に相手方の情報を利用するピア・ツー・ピアシステムを構築した場合である。

【 0 1 4 7 】

図 2 0 において、ステップ S 1 で本発明の情報収集装置 1 0 は、例えば他の情報収集装置 1 0 - 1 と通信して、他の情報収集装置 1 0 - 1 に登録しているイベント収集先サイトの情報を取得する。

【 0 1 4 8 】

この他の情報収集装置 1 0 - 1 から取得してイベント収集先サイトにつき、自分のイベント収集先サイト登録部 1 8 の内容と比較し、イベント収集先サイトが

異なればステップ S 3 で他の情報収集装置 1 0 - 1 のイベント収集先サイトの方が良いか否か判断する。

【 0 1 4 9 】

このステップ S 3 におけるイベント収集先サイトの判定条件としては、情報登録の早さを示す情報取得日時、ドキュメントのバイト数などの情報量に基づいてイベント収集先サイトの良し悪しを数値化し、他の情報収集装置 1 0 - 1 で同じく数値化している値と比較して良い方を採用し、ステップ S 4 で採用した他の情報収集装置 1 0 - 1 から収集したイベント収集先サイトを自分自身のイベント収集先サイト登録部 1 8 に登録する。

【 0 1 5 0 】

またステップ S 5 で他の情報収集装置 1 0 - 1 と通信して、情報収集先サイトの登録情報を取得し、ステップ S 6 で自分自身の情報収集先サイト登録部 2 0 の登録サイトと異なれば、ステップ S 7 でステップ S 3 の収集先サイトの場合と同様にして数値化した値の比較により、他の情報収集装置 1 0 - 1 の情報収集先サイトの良し悪しを判断し、良ければステップ S 8 で自分自身の情報収集先サイト登録部 2 0 に取得した情報収集先サイトを登録する。

【 0 1 5 1 】

これに続くステップ S 9 ～ステップ S 1 7 の処理は、図 4 のステップ S 4 ～ S 1 1 の処理と同じである。

【 0 1 5 2 】

図 2 2 は、図 2 の情報収集装置 1 0 におけるキーワード抽出部 2 4 におけるキーワード抽出処理の詳細を示したフローチャートである。

【 0 1 5 3 】

図 2 2 において、キーワード抽出処理は、まずステップ S 1 でイベント収集先サイトから得られたドキュメントの変化領域、例えば図 5 (B) の情報 4 0 の文章を形態素解析して品詞に分解する。ここでイベント収集先サイトから得られた変化領域の文章は、製品名やウィルス名などのイベントを特定する固有名詞を含んでいることから、形態素解析したデータからステップ S 2 で固有名詞のみを抽出する。

【0154】

続いてステップS3でキーワードデータベース30内の固有名詞と比較し、キーワードデータベース30内にあるか否かチェックする。キーワードデータベース30に無ければ、ステップS2で抽出した固有名詞をステップS4でキーワードとして保持する。一方、ステップS3でキーワードデータベース30に登録されている場合には、この固有名詞はすでにキーワードとして使われていることからステップS5で固有名詞を破棄する。

【0155】

このようなステップS1～S5の処理をステップS6で変化領域の文章内の全固有名詞について終了するまで繰り返す。ステップS6において全固有名詞について終了が判別されると、ステップS7においてステップS4で保持された固有名詞をキーワードデータベース30に登録して更新した後、ステップS8で保持した固有名詞をキーワードとして情報検索部26に出力する。

【0156】

ここで図22のキーワード抽出処理において、イベント収集先サイトから取得したドキュメントの変化領域の文章から複数のキーワードを抽出した場合には、それらのキーワードの重要度を判断して優先順位をつけ、この優先順位付きのキーワードを情報検索部26に出力し、優先順位に従ったキーワードの使用で情報検索を行うようにしても良い。

【0157】

複数のキーワードを抽出した時の重要度を判断した優先順位の付け方としては

- (1) 外部リンクが設定されているキーワード
- (2) 外部リンク先ドキュメント内で登場回数の多いキーワード
- (3) 「」や“”など特定の記号で囲まれているキーワード
- (4) ボールド、、赤文字、などの強調指定されているキーワード

を抽出し、ドキュメントの抽出内容に応じて固有のポイントをつけてその総和を求める。例えば(1)・(2)は一個につき3点、(3)は10点などとして合

計点数を求め、合計点数の高い順にキーワードに優先順位をつける。

【0158】

図23及び図24は、図2の情報収集装置10における情報収集装置の他の実施形態のフローチャートであり、この実施形態にあつては、イベント収集先サイトから得られたドキュメントに追加された新規情報及び削除情報の履歴を保存することで、どのような時系列でイベント収集先の情報が更新され、また削除されたかをユーザーがわかるようにしたことを特徴とする。

【0159】

図23において、ステップS1～S6の処理は、図4のステップS1～S6と同じである。ステップS6でイベント収集先サイトのドキュメントをリファレンスと比較して変化があつた場合には、ステップS7で削除無しの新規情報の追加更新か否かチェックする。

【0160】

ここでイベント収集先サイトのドキュメントの更新には図5(A)(B)のように、古い情報38を破棄した新規な情報40を追加する更新の形態と、図6のように古い情報38を残してまま新規な情報40を追加する更新形態の2つがある。

【0161】

このためステップS7で図6の削除無しの新規情報の追加更新を判別した場合には、ステップS8で例えば図6(B)の取得したデータであるドキュメント36-12の変化領域となる新規な情報40を抽出し、これを変化領域情報履歴に加えて更新する。

【0162】

一方、ステップS7で図5のような削除ありの新規情報の追加更新を判別した場合には、図5(A)のリファレンスとしてのドキュメント36-1と図5(B)の新たに得られたドキュメント36-2を比較し、変化領域としてドキュメント36-1の破棄領域となる情報38とドキュメント36-2の追加領域となる新規な情報40を抽出する。

【0163】

そしてステップ S 11 で新規情報履歴に追加された新規な情報 40 を加えて更新し、またステップ S 12 で削除情報履歴に破棄領域となる削除した情報 38 加えて更新する。このように更新された新規情報履歴および削除情報履歴は、ユーザが必要に応じて参照することができ、それぞれの履歴が時系列に従った一覧として表示される。

【0164】

このようなステップ S 7～S 9、またはステップ S 7～S 12 の更新の履歴処理が済んだならば、ステップ S 13 でイベント収集先保存ページとしてのリファレンスを新たに比較したドキュメントで更新し、図 24 のステップ S 14～S 16 において、イベント収集先サイトの変化領域からイベントを特定するキーワードを抽出し、キーワードより情報収集先サイトのドキュメントを検索し、ユーザーに通知する。

【0165】

図 25 及び図 26 は、イベント収集先サイトについて更新された情報一覧の履歴保存につき、更新領域からキーワードを抽出して、キーワードによる更新履歴を保存利用できるようにしたことを特徴とする。

【0166】

図 25 及び図 26 において、ステップ S 1～S 7、ステップ S 9、S 11～S 16 の処理は図 23、図 24 のフローチャートと同じである。これに対し、図 25 のステップ 8、S 10 において、イベント収集先サイトから取得したデータ、すなわちドキュメントの変化領域からキーワードを抽出している。

【0167】

即ち、ステップ S 8 にあっては、ステップ S 7 で判別された図 6 (B) の変化領域の情報 40 の文章から、例えばキーワードとして「XXX」を抽出し、このキーワード「XXX」をステップ S 9 で新規情報履歴に追加して更新する。またステップ S 7 で図 5 のような削除更新を判別した場合には、ステップ S 12 に進み図 5 (A) の破棄領域として削除される情報 38 からキーワード「AAA」を抽出し、また図 5 (B) の追加領域となる情報 40 からキーワード「XXX」を抽出し、ステップ S 11 でキーワード「XXX」を新規情報履歴に追加して更新

し、またステップ S 1 2 でキーワード「A A A」を削除情報履歴に追加して更新する。

【 0 1 6 8 】

このようにイベント収集先サイトのドキュメントの新規情報履歴及び削除情報履歴をキーワードの一覧表として保存利用できることで、新規情報履歴及び削除情報履歴をユーザーが読み出によりキーワード一覧化として表示され、新製品の時系列的な更新の様子を容易に把握することができる。

【 0 1 6 9 】

図 2 7 及び図 2 8 は、図 2 の情報収集装置 1 0 における本発明の情報収集処理の他の情報収集装置のフローチャートであり、この実施形態にあってはイベント収集先サイトの更新で得られた変化領域に存在するリンク先からドキュメントをダウンロードして保存したことを特徴とする。

【 0 1 7 0 】

図 2 7 及び図 2 8 のフローチャートにおけるステップ S 1 ～ S 8 及びステップ S 1 0、S 1 1、S 1 3 ～ S 1 8 の処理は、図 2 5、図 2 6 におけるステップ S 1 ～ S 8 及びステップ S 9 ～ S 1 6 の処理と同じである。図 2 7 において、新たにステップ S 9、S 1 2 の処理が加えられている。

【 0 1 7 1 】

ステップ S 9 の処理は、ステップ S 7 における図 6 のようなイベント収集先サイトからダウンロードした変化領域となる新規な情報 4 0 の中に他のサイトのリンク情報が含まれている場合、このリンク情報により他のサイトにアクセスして変化領域に示されたリンク先のドキュメントをダウンロードしてドキュメント保存部 3 2 に保存する。

【 0 1 7 2 】

ステップ S 1 2 の処理は、ステップ S 7 における図 5 のようなイベント収集先サイトからダウンロードした変化領域となる新規な情報 4 0 の中に他のサイトのリンク情報が含まれている場合、このリンク情報により他のサイトにアクセスして変化領域に示されたリンク先のドキュメントをダウンロードしてドキュメント保存部 3 2 に保存する。

【0173】

これによってイベント収集先サイトの更新により更新履歴のリンク情報が削除されても、削除されたリンク先のサーバからドキュメントを保存しているためユーザーは新規情報履歴を見た際のリンク先としてドキュメント保存部32からすでに削除されているリンク先サーバからのドキュメントにアクセスすることができる。

【0174】

尚、上記の実施形態は情報収集装置10として、例えば図3のようなハードウェア資源をもつパーソナルコンピュータへの適用を例にとるものであったが、これ以外に携帯端末や適宜のコンピュータ装置にそのまま適用することができる。また本発明はその目的と利点を損ねることのない適宜の変形を含み、更に上記の実施形態に示した数値による限定は受けない。

【0175】

(付記)

(付記1)

ネットワークに接続するネットワーク接続部と、
ネットワークあるいは実社会上で発生したイベントの有無を検出するためのイベント収集先サイトを登録するイベント収集先サイト登録部と、
テキスト、画像、音声等のデータを含むドキュメントを収集するための情報収集先サイトを登録する情報収集先サイト登録部と、
前記登録されたイベント収集先サイトから情報を取得し、取得情報の更新の有無からイベント発生の有無を検出するイベント検出部と、
前記イベント検出部で検出された更新領域からキーワードを抽出するキーワード抽出部と、
前記キーワード抽出部で抽出されたキーワードを用いて前記登録された情報収集先サイト内のドキュメントを検索する情報検索部と、
前記情報検索部による検索結果をユーザに通知する情報通知部と、
を備えたことを特徴とする情報収集装置。(1)

【0176】

(付記2)

付記1記載の情報収集装置に於いて、前記イベント検出部は、前記イベント収集先サイトをアクセスしてサイト内のドキュメントをダウンロードしてリファレンスとして保存し、その後、前記イベント収集先サイトからダウンロードしたドキュメントと前記リファレンスの比較により更新の有無からイベント発生の有無を検出すると共に、該ダウンロードドキュメントを用いて前記リファレンスを更新することを特徴とする情報収集装置。

【0177】

(付記3)

付記1記載の情報収集装置に於いて、前記情報検索部は、前記情報収集先サイトをアクセスしてサイト内のドキュメントをダウンロードし、ダウンロードしたドキュメントの中から前記キーワードを用いて該当するドキュメント部分を検索することを特徴とする情報収集装置。

【0178】

(付記4)

付記1記載の情報収集装置に於いて、更に、前記情報検索部により前記情報収集先サイトから取得されたドキュメントを保存するドキュメント保存部を設けたことを特徴とする情報収集装置。

【0179】

(付記5)

付記1記載の情報収集装置に於いて、前記情報検索部は、前記キーワード抽出部で抽出されたキーワードを用いて一定期間に亘り定期的に前記登録された情報収集先サイト内のドキュメントを検索することを特徴とする情報収集装置。

【0180】

(付記6)

付記1記載の情報収集装置に於いて、
前記イベント収集先サイト登録部は、ネットワークを介してイベント収集先リストサーバからイベント収集先サイトを取得して登録し、

前記情報収集先サイト登録部は、ネットワークを介して情報収集先リストサーバから情報収集先サイトを取得して登録する、
ことを特徴とする情報収集装置。

【0181】

(付記7)

付記1記載の情報収集装置に於いて、
前記イベント収集先サイト登録部は、ネットワークを介して同一構成を備えた他の情報収集装置からイベント収集先サイトを取得して登録し、
前記情報収集先サイト登録部は、ネットワークを介して同一構成を備えた情報収集装置から情報収集先サイトを取得して登録する、
ことを特徴とする情報収集装置。

【0182】

(付記8)

付記1記載の情報収集装置に於いて、前記キーワード抽出部は、前記イベント検出部で検出された更新領域を形態素解析して品詞ごとに分割した後に固有名詞のみを抽出し、抽出した名詞がキーワードデータベースに登録されている既存のキーワードと異なる場合に、前記情報検索部に抽出した固有名詞をキーワードとして出力することを特徴とする情報収集装置。

【0183】

(付記9)

付記1記載の情報収集装置に於いて、前記イベント検出部は、イベント発生を検出したイベント収集先サイトの更新領域に新規情報の追加のみがあった場合は、前記新規情報の履歴を保存し、前記更新領域に新規情報が追加されると同時に古い情報が削除された場合は、前記新規情報の履歴と前記削除情報の履歴を保存し、前記情報通知部により保存履歴をユーザに通知可能としたことを特徴とする情報収集装置。

【0184】

(付記10)

付記1記載の情報収集装置に於いて、前記イベント検出部は、イベント発生を

検出したイベント収集先サイトの更新領域に新規情報の追加のみが合った場合は、前記新規情報の履歴として前記キーワード抽出部で抽出されたキーワードを保存し、前記更新領域に新規情報が追加されると同時に古い情報が削除された場合は、新規情報の履歴と削除情報の履歴として前記キーワード抽出部で抽出されたキーワードを保存し、前記情報通知部により保存履歴としての前記キーワードをユーザに通知可能としたことを特徴とする情報収集装置。

【 0 1 8 5 】

(付記 1 1)

イベント収集先サイト登録部により、ネットワークあるいは実社会上で発生したイベントの有無を検出するためのイベント収集先サイトを登録するイベント収集先サイト登録ステップと、

情報収集先サイト登録部により、テキスト、画像、音声等のデータを含むドキュメントを収集するための情報収集先サイトを登録する情報収集先サイト登録ステップと、

イベント検出部により、前記登録されたイベント収集先サイトから情報を取得し、取得情報の更新の有無からイベント発生の有無を検出するイベント検出ステップと、

キーワード抽出部により、前記イベント検出ステップで検出された更新領域からキーワードを抽出するキーワード抽出ステップと、

情報検索部により、前記キーワード抽出部で抽出されたキーワードを用いて前記登録された情報収集先サイト内のドキュメントを検索する情報検索ステップと、
情報通知部により、前記情報検索ステップの検索結果をユーザに通知する情報通知ステップと、

を備えたことを特徴とする情報収集方法。(2)

【 0 1 8 6 】

(付記 1 2)

付記 1 1 記載の情報収集方法に於いて、前記イベント検出ステップは、前記イベント収集先サイトをアクセスしてサイト内のドキュメントをダウンロードしてリファレンスとして保存し、その後、前記イベント収集先サイトからダウンロー

ドしたドキュメントと前記リファレンスの比較により更新の有無からイベント発生の有無を検出することを特徴とする情報収集方法。(3)

【0187】

(付記13)

付記11記載の情報収集方法に於いて、前記情報検索ステップは、前記情報収集先サイトをアクセスしてサイト内のドキュメントをダウンロードし、ダウンロードしたドキュメントの中から前記キーワードを用いて該当するドキュメント部分を検索することを特徴とする情報収集方法。

【0188】

(付記14)

付記11記載の情報収集方法に於いて、更に、前記情報検索ステップにより前記情報収集先サイトから取得されたドキュメントをドキュメント保存部に保存するドキュメント保存ステップを設けたことを特徴とする情報収集方法。(4)

【0189】

(付記15)

付記11記載の情報収集方法に於いて、前記情報検索ステップは、前記キーワードを用いたドキュメントの検索回数を計数し、一定期間経過した時のドキュメントの検索回数が予め定めた閾値を超えた場合は前記キーワードによるドキュメントの情報検索を再度一定期間継続し、前記閾値以下であった場合は、前記キーワードによる情報検索を停止することを特徴とする情報収集方法。

【0190】

(付記16)

付記11記載の情報収集方法に於いて、
前記イベント収集先サイト登録ステップは、ネットワークを介してイベント収集先リストサーバからイベント収集先サイトを取得して登録し、
前記情報収集先サイト登録ステップは、ネットワークを介して情報収集先リストサーバから情報収集先サイトを取得して登録する、
ことを特徴とする情報収集方法。

【0191】

(付記 17)

付記 11 記載の情報収集方法に於いて、

前記イベント収集先サイト登録ステップは、ネットワークを介して同一構成を備えた他の情報収集装置からイベント収集先サイトを取得して登録し、

前記情報収集先サイト登録ステップは、ネットワークを介して同一構成を備えた情報収集装置から情報収集先サイトを取得して登録する、

ことを特徴とする情報収集方法。

【0192】

(付記 18)

付記 11 記載の情報収集方法に於いて、前記キーワード抽出ステップは、前記イベント検出ステップで検出された更新領域を形態素解析して品詞ごとに分割した後に固有名詞のみを抽出し、抽出した名詞がキーワードデータベースに登録されている既存のキーワードと異なる場合に、前記情報検索ステップに抽出した固有名詞をキーワードとして出力することを特徴とする情報収集方法。

【0193】

(付記 19)

付記 11 記載の情報収集方法に於いて、前記イベント検出ステップは、イベント発生を検出したイベント収集先サイトの更新領域に新規情報の追加のみがあった場合は、前記新規情報の履歴を保存し、前記更新領域に新規情報が追加されると同時に古い情報が削除された場合は、前記新規情報の履歴と前記削除情報の履歴を保存し、前記情報通知部により保存履歴をユーザに通知可能としたことを特徴とする情報収集方法。

【0194】

(付記 20)

付記 11 記載の情報収集方法に於いて、前記イベント検出ステップは、イベント発生を検出したイベント収集先サイトの更新領域に新規情報の追加のみがあった場合は、前記新規情報の履歴として前記キーワード抽出部で抽出されたキーワードを保存し、前記更新領域に新規情報が追加されると同時に古い情報が削除された場合は、新規情報の履歴と削除情報の履歴として前記キーワード抽出部で抽

出されたキーワードを保存し、前記情報通知部により保存履歴としての前記キーワードをユーザに通知可能としたことを特徴とする情報収集方法。

【0195】

(付記21)

コンピュータに、
ネットワークあるいは実社会上で発生したイベントの有無を検出するためのイベント収集先サイトを登録するイベント収集先サイト登録ステップと、
テキスト、画像、音声等のデータを含むドキュメントを収集するための情報収集先サイトを登録する情報収集先サイト登録ステップと、
前記登録されたイベント収集先サイトから情報を取得し、取得情報の更新の有無からイベント発生の有無を検出するイベント検出ステップと、
前記イベント検出ステップで検出された更新領域からキーワードを抽出するキーワード抽出ステップと、
前記キーワード抽出部で抽出されたキーワードを用いて前記登録された情報収集先サイト内のドキュメントを検索する情報検索ステップと、
前記情報検索ステップの検索結果をユーザに通知する情報通知ステップと、
を実行させることを特徴とするプログラム。(5)

【0196】

(付記22)

付記21記載のプログラムに於いて、前記イベント検出ステップは、前記イベント収集先サイトをアクセスしてサイト内のドキュメントをダウンロードしてリファレンスとして保存し、その後、前記イベント収集先サイトからダウンロードしたドキュメントと前記リファレンスの比較により更新の有無からイベント発生の有無を検出することを特徴とするプログラム。

【0197】

(付記23)

付記21記載のプログラムに於いて、前記情報検索ステップは、前記情報収集先サイトをアクセスしてサイト内のドキュメントをダウンロードし、ダウンロードしたドキュメントの中から前記キーワードを用いて該当するドキュメント部分

を検索することを特徴とするプログラム。

【 0 1 9 8 】

(付記 2 4)

付記 2 1 記載のプログラムに於いて、更に、前記情報検索ステップにより前記情報収集先サイトから取得されたドキュメントをドキュメント保存部に保存するドキュメント保存ステップを設けたことを特徴とするプログラム。

【 0 1 9 9 】

(付記 2 5)

付記 2 1 記載のプログラムに於いて、前記情報検索ステップは、前記キーワード抽出ステップで抽出されたキーワードを用いて一定期間に亘り定期的に前記登録された情報収集先サイト内のドキュメントを検索することを特徴とするプログラム。

【 0 2 0 0 】

(付記 2 6)

付記 2 1 記載のプログラムに於いて、

前記イベント収集先サイト登録ステップは、ネットワークを介してイベント収集先リストサーバからイベント収集先サイトを取得して登録し、

前記情報収集先サイト登録ステップは、ネットワークを介して情報収集先リストサーバから情報収集先サイトを取得して登録する、
ことを特徴とするプログラム。

【 0 2 0 1 】

(付記 2 7)

付記 2 1 記載のプログラムに於いて、

前記イベント収集先サイト登録ステップは、ネットワークを介して同一構成を備えた他の情報収集装置からイベント収集先サイトを取得して登録し、

前記情報収集先サイト登録ステップは、ネットワークを介して同一構成を備えた情報収集装置から情報収集先サイトを取得して登録する、
ことを特徴とするプログラム。

【 0 2 0 2 】

(付記 2 8)

付記 2 1 記載のプログラムに於いて、前記キーワード抽出ステップは、前記イベント検出ステップで検出された更新領域を形態素解析して品詞ごとに分割した後に固有名詞のみを抽出し、抽出した名詞がキーワードデータベースに登録されている既存のキーワードと異なる場合に、前記情報検索ステップに抽出した固有名詞をキーワードとして出力することを特徴とするプログラム。

【 0 2 0 3 】

(付記 2 9)

付記 2 1 記載のプログラムに於いて、前記イベント検出ステップは、イベント発生を検出したイベント収集先サイトの更新領域に新規情報の追加のみがあった場合は、前記新規情報の履歴を保存し、前記更新領域に新規情報が追加されると同時に古い情報が削除された場合は、前記新規情報の履歴と前記削除情報の履歴を保存し、前記情報通知部により保存履歴をユーザに通知可能としたことを特徴とするプログラム。

【 0 2 0 4 】

(付記 3 0)

付記 2 1 記載のプログラムに於いて、前記イベント検出ステップは、イベント発生を検出したイベント収集先サイトの更新領域に新規情報の追加のみが合った場合は、前記新規情報の履歴として前記キーワード抽出部で抽出されたキーワードを保存し、前記更新領域に新規情報が追加されると同時に古い情報が削除された場合は、新規情報の履歴と削除情報の履歴として前記キーワード抽出部で抽出されたキーワードを保存し、前記情報通知部により保存履歴としての前記キーワードをユーザに通知可能としたことを特徴とするプログラム。

【 0 2 0 5 】

更に上記の付記に対応して本発明は次に列挙する実施例をもつ。

【 0 2 0 6 】

(実施例 1)

付記 4 記載の情報収集装置に於いて、前記ドキュメント保存部は、前記情報検

索部で検索された検索済みドキュメントを、検索に使用したキーワードをインデックスとして保存することを特徴とする。

【 0 2 0 7 】

(実施例 2)

付記 4 記載の情報収集装置に於いて、前記情報検索部は、前記情報収集サイト先を定期的にアクセスしてサイト内のドキュメントをダウンロードして前記ドキュメント保存部に保存し、前記キーワード抽出部で抽出されたキーワードを用いて前記ドキュメント保存部に保存されたドキュメントを検索することを特徴とする。

【 0 2 0 8 】

(実施例 3)

実施例 2 記載の情報収集装置に於いて、前記情報検索部は、検索された回数をドキュメントごとに計数し、検索回数が予め定めた閾値以下のドキュメントを前記ドキュメント保存部から削除することを特徴とする。

【 0 2 0 9 】

(実施例 4)

実施例 3 記載の情報収集装置に於いて、前記情報検索部は、前記ドキュメント保存部の空き容量の不足を判断した場合、前記前記閾値を大きくして閾値以下の検索回数のドキュメントを前記ドキュメント保存部から削除することを特徴とする。

【 0 2 1 0 】

(実施例 5)

実施例 2 記載の情報収集装置に於いて、前記イベント検出部は、前記イベント収集先サイトから取得したドキュメントの更新領域に加え、削除された破棄領域を検出し、前記キーワード抽出部により前記破棄領域のキーワードを抽出して前記ドキュメント保存部に保存されたドキュメントを検索して削除することを特徴とする。

【 0 2 1 1 】

(実施例 6)

付記 5 記載の情報収集装置に於いて、前記情報検索部は、前記キーワードを用いたドキュメントの検索回数を計数し、一定期間経過した時のドキュメントの検索回数が予め定めた閾値を超えた場合は前記キーワードによるドキュメントの情報検索を再度一定期間継続し、前記閾値以下であった場合は、前記キーワードによる情報検索を停止することを特徴とする。

【 0 2 1 2 】

(実施例 7)

付記 8 記載の情報処理装置に於いて、前記キーワード抽出部は、前記情報検索部にキーワードとして出力した固有名詞を前記キーワードデータベースに追加登録することを特徴とする。

【 0 2 1 3 】

(実施例 8)

付記 8 記載の情報収集装置に於いて、前記キーワード抽出部は、前記イベント検出部で検出された情報の更新領域から複数のキーワードを抽出した場合、前記更新領域の内容に基づいて各キーワードの優先順位を付与して前記情報検索部に出力することを特徴とする。

【 0 2 1 4 】

(実施例 9)

付記 9 記載の情報収集装置に於いて、前記イベント検出部は、更新領域に追加された新規情報に外部サイトのリンクがあった場合、前記外部リンク先のドキュメントをダウンロードして前記ドキュメント保存部に保存すると共に前記新規情報情報の履歴に前記ドキュメント保存部の保存ドキュメントをリンクさせることを特徴とする。

【 0 2 1 5 】

(実施例 1 0)

付記 1 4 記載の情報収集方法に於いて、前記ドキュメント保存ステップは、前記情報検索ステップで検索された検索済みドキュメントを、検索に使用したキーワードをインデックスとして保存することを特徴とする。

【 0 2 1 6 】

(実施例 1 1)

付記 1 4 記載の情報収集方法に於いて、前記情報検索ステップは、前記情報収集先サイトを定期的にアクセスしてサイト内のドキュメントをダウンロードして前記ドキュメント保存部に保存し、前記キーワード抽出ステップで抽出されたキーワードを用いて前記ドキュメント保存部に保存されたドキュメントを検索することを特徴とする。

【 0 2 1 7 】

(実施例 1 2)

実施例 1 1 記載の情報収集方法に於いて、前記情報検索ステップは、検索された回数をドキュメントごとに計数し、検索回数が予め定めた閾値以下のドキュメントを前記ドキュメント保存部から削除することを特徴とする。

【 0 2 1 8 】

(実施例 1 3)

実施例 1 2 記載の情報収集方法に於いて、前記情報検索ステップは、前記ドキュメント保存部の空き容量の不足を判断した場合、前記前記閾値を大きくして閾値以下の検索回数のドキュメントを前記ドキュメント保存部から削除することを特徴とする。

【 0 2 1 9 】

(実施例 1 4)

実施例 1 1 記載の情報収集方法に於いて、前記イベントステップ、前記イベント収集先サイトから取得したドキュメントの更新領域に加え、削除された破棄領域を検出し、前記破棄領域のキーワードを抽出して前記ドキュメント保存部に保存されたドキュメントを検索して削除することを特徴とする。

【 0 2 2 0 】

(実施例 1 5)

付記 1 5 記載の情報収集方法に於いて、前記情報検索ステップは、前記キーワードを用いたドキュメントの検索回数を計数し、一定期間経過した時のドキュメントの検索回数が予め定めた閾値を超えた場合は前記キーワードによるドキュメントの情報検索を再度一定期間継続し、前記閾値以下であった場合は、前記キ

ワードによる情報検索を停止することを特徴とする。

【 0 2 2 1 】

(実施例 1 6)

付記 1 8 記載の情報処理方法に於いて、前記キーワード抽出ステップは、前記情報検索ステップにキーワードとして出力した固有名詞を前記キーワードデータベースに追加登録することを特徴とする。

【 0 2 2 2 】

(実施例 1 7)

付記 1 8 記載の情報収集方法に於いて、前記キーワード抽出ステップは、前記イベント検出ステップで検出された情報の更新領域から複数のキーワードを抽出した場合、前記更新領域の内容に基づいて各キーワードの優先順位を付与して前記情報検索部に出力することを特徴とする。

【 0 2 2 3 】

(実施例 1 8)

付記 1 9 記載の情報収集方法に於いて、前記イベント検出ステップは、更新領域に追加された新規情報に外部サイトのリンクがあった場合、前記外部リンク先のドキュメントをダウンロードして前記ドキュメント保存部に保存すると共に前記新規情報情報の履歴に前記ドキュメント保存部の保存ドキュメントをリンクさせることを特徴とする。

【 0 2 2 4 】

(実施例 1 9)

付記 2 4 記載のプログラムに於いて、前記ドキュメント保存ステップは、前記情報検索ステップで検索された検索済みドキュメントを、検索に使用したキーワードをインデックスとして保存することを特徴とする。

【 0 2 2 5 】

(実施例 2 0)

付記 2 4 記載のプログラムに於いて、前記情報検索ステップは、前記情報収集サイト先を定期的にアクセスしてサイト内のドキュメントをダウンロードして前記ドキュメント保存部に保存し、前記キーワード抽出ステップで抽出されたキー

ワードを用いて前記ドキュメント保存部に保存されたドキュメントを検索することを特徴とする。

【 0 2 2 6 】

(実施例 2 1)

実施例 2 0 記載のプログラムに於いて、前記情報検索ステップは、検索された回数をドキュメントごとに計数し、検索回数が予め定めた閾値以下のドキュメントを前記ドキュメント保存部から削除することを特徴とする。

【 0 2 2 7 】

(実施例 2 2)

実施例 2 1 記載のプログラムに於いて、前記情報検索ステップは、前記ドキュメント保存部の空き容量の不足を判断した場合、前記前記閾値を大きくして閾値以下の検索回数のドキュメントを前記ドキュメント保存部から削除することを特徴とする。

【 0 2 2 8 】

(実施例 2 3)

実施例 2 0 記載のプログラムに於いて、前記イベントステップ、前記イベント収集先サイトから取得したドキュメントの更新領域に加え、削除された破棄領域を検出し、前記破棄領域のキーワードを抽出して前記ドキュメント保存部に保存されたドキュメントを検索して削除することを特徴とする。

【 0 2 2 9 】

(実施例 2 4)

付記 2 5 記載のプログラムに於いて、前記情報検索ステップは、前記キーワードを用いたドキュメントの検索回数を計数し、一定期間経過した時のドキュメントの検索回数が予め定めた閾値を超えた場合は前記キーワードによるドキュメントの情報検索を再度一定期間継続し、前記閾値以下であった場合は、前記キーワードによる情報検索を停止することを特徴とする。

【 0 2 3 0 】

(実施例 2 5)

付記 2 8 記載のプログラムに於いて、前記キーワード抽出ステップは、前記情

報検索ステップにキーワードとして出力した固有名詞を前記キーワードデータベースに追加登録することを特徴とする。

【 0 2 3 1 】

(実施例 2 6)

付記 2 8 記載のプログラムに於いて、前記キーワード抽出ステップは、前記イベント検出ステップで検出された情報の更新領域から複数のキーワードを抽出した場合、前記更新領域の内容に基づいて各キーワードの優先順位を付与して前記情報検索部に出力すること特徴とする。

【 0 2 3 2 】

(実施例 2 7)

付記 2 9 記載のプログラムに於いて、前記イベント検出ステップは、更新領域に追加された新規情報に外部サイトのリンクがあった場合、前記外部リンク先のドキュメントをダウンロードして前記ドキュメント保存部に保存すると共に前記新規情報情報の履歴に前記ドキュメント保存部の保存ドキュメントをリンクさせることを特徴とする。

【 0 2 3 3 】

【発明の効果】

以上説明してきたように本発明によれば、イベント収集先サイトとして特定のサイトを監視し、サイト情報更新によるイベント発生を検出した場合にはその更新内容から新製品発表、新ウィルス発生などといったイベント特定するキーワードを抽出し、抽出したキーワードを使用して情報収集先サイトから情報を検索してユーザに表示し、これによってユーザがキーワードなどの情報を特定するための単語を設定する必要がなく、ユーザにとって未知の情報であっても複数の情報提供先から有効な情報を自動的に収集して、ユーザに通知することができる。

【 0 2 3 4 】

特に情報の迅速な収集を必要とする新製品情報や新ウィルス発生情報などにつき、ユーザはイベント収集先サイトを予め登録しておくだけで新製品発表や新ウィルス発生というイベント発生の通知とともに、新製品の内容、評判、価格など

といった情報、また新ウィルス発生についてはパソコンメーカーによるウィルス対策の情報をユーザに通知でき、ネットワーク上で起きているダイナミックイベントに対し、迅速かつ適切に必要な情報を収集してユーザに与えることができる。

【図面の簡単な説明】

【図 1】

本発明の原理構成の説明図

【図 2】

本発明による情報収集装置の実施形態の機能ブロック図

【図 3】

図 2 の実施形態が適用されるコンピュータのハードウェア資源の説明図

【図 4】

図 2 の実施形態による情報収集処理の基本的処理手順のフローチャート

【図 5】

イベント収集先サイトから取得した新製品発売情報の説明図

【図 6】

イベント収集先サイトから取得した新製品発売情報の他の形態の説明図

【図 7】

情報収集先サイトからキーワードにより検索したドキュメントを保存する本発明の他の実施形態のフローチャート

【図 8】

情報収集先サイトから収集したドキュメントを保存した後にキーワードにより保存ドキュメントを検索する本発明の他の実施形態のフローチャート

【図 9】

検索回数の少ない保存ドキュメントを削除する本発明の他の実施形態のフローチャート

【図 1 0】

図 9 に続く本発明の他の実施形態のフローチャート

【図 1 1】

保存ドキュメントを削除する検索回数の閾値を大きくして十分な空き容量を確保する本発明の他の実施形態のフローチャート

【図 1 2】

図 1 1 に続く本発明の他の実施形態のフローチャート

【図 1 3】

イベント収集先サイトの情報更新で削除された破棄領域からキーワードを抽出して保存ドキュメントを削除する本発明の他の実施形態のフローチャート

【図 1 4】

図 1 3 に続く本発明の他の実施形態のフローチャート

【図 1 5】

イベント発生検出から一定時間経過するまで定期的にキーワードによりドキュメントを検索する本発明の他の実施形態のフローチャート

【図 1 6】

一定期間のあいだ検索回数が閾値以下であれば発生イベントの新規性が失われたとして情報収集を停止する本発明の他の実施形態のフローチャート

【図 1 7】

図 1 6 に続く本発明の他の実施形態のフローチャート

【図 1 8】

リストサーバからイベント収集サイト及び情報収集先サイトを取得する本発明の他の実施形態のフローチャート

【図 1 9】

図 1 8 に続く本発明の他の実施形態のフローチャート

【図 2 0】

他の情報収集先装置からイベント収集サイト及び情報収集先サイトを取得する本発明の他の実施形態のフローチャート

【図 2 1】

図 2 0 に続く本発明の他の実施形態のフローチャート

【図 2 2】

本発明におけるキーワード抽出処理のフローチャート

【図 2 3】

イベント収集先サイトの更新に伴う新規情報と削除情報の履歴を保存して利用する本発明の他の実施形態のフローチャート

【図 2 4】

図 2 3 に続く本発明の他の実施形態のフローチャート

【図 2 5】

イベント収集先サイトの更新に伴う新規情報と削除情報の履歴をキーワードとして保存して利用する本発明の他の実施形態のフローチャート

【図 2 6】

図 2 5 に続く本発明の他の実施形態のフローチャート

【図 2 7】

イベント収集先サイトの更新に伴う新規情報にある外部リンク先からドキュメントを取得して保存する本発明の他の実施形態のフローチャート

【図 2 8】

図 2 7 に続く本発明の他の実施形態のフローチャート

【符号の説明】

- 1 0, 1 0 - 1, 1 0 - 2 : 情報収集装置
- 1 1 : インターネット
- 1 2 - 1 ~ 1 2 - 3 : イベント収集先サイト
- 1 4 - 1 ~ 1 4 - 3 : 情報収集先サイト
- 1 5 - 1 : 情報収集先リストサーバ
- 1 5 - 2 : イベント収集先リストサーバ
- 1 6 : ネットワーク接続部
- 1 8 : イベント収集先サイト登録部
- 2 0 : 情報収集先サイト登録部
- 2 2 : イベント検出部
- 2 4 : キーワード抽出部

26 : 情報検索部

28 : 情報通知部

30 : キーワードデータベース

32 : ドキュメント保存部

34 : 表示部

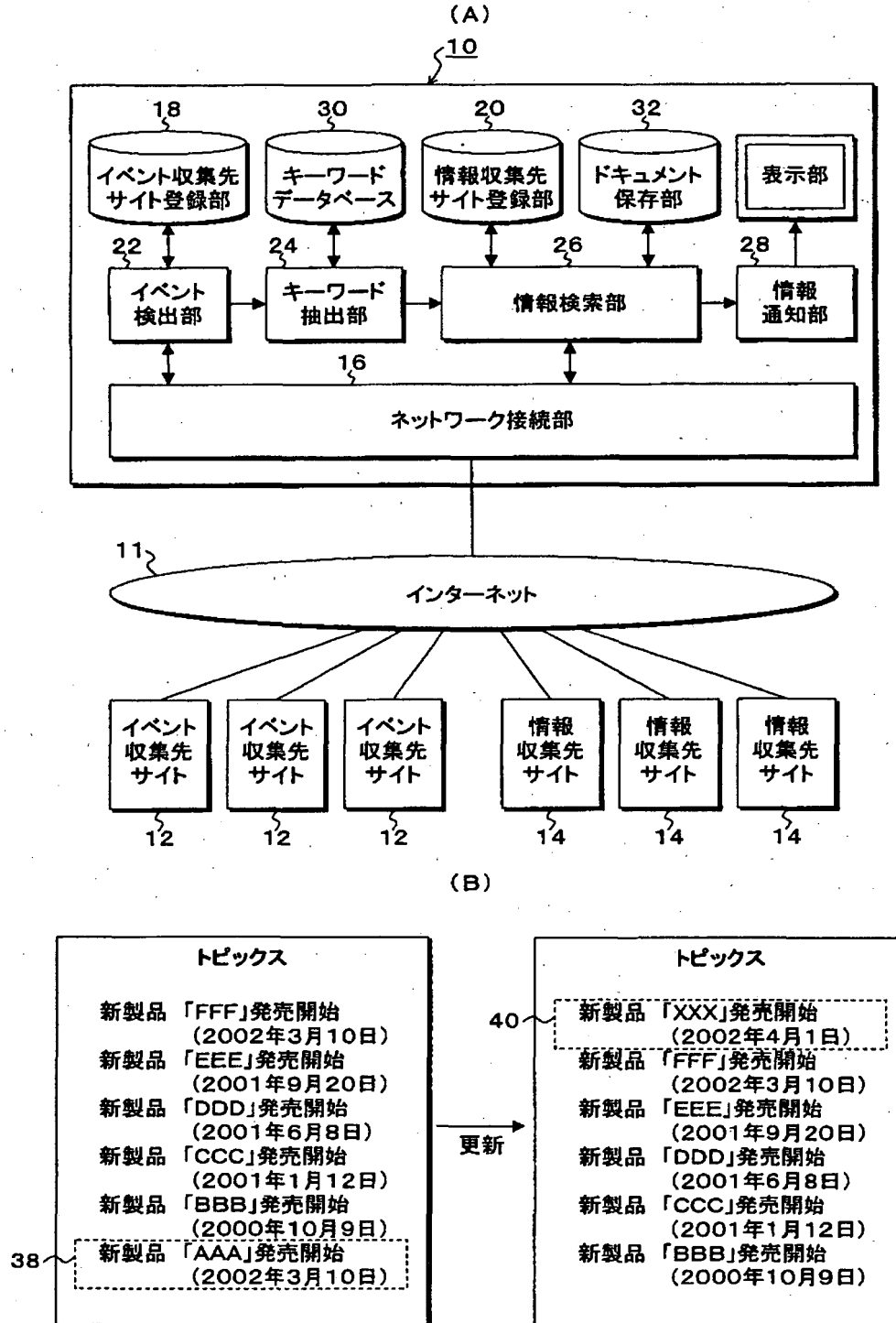
36-1、36-2 : ドキュメント

38, 40 : 情報

【書類名】 図面

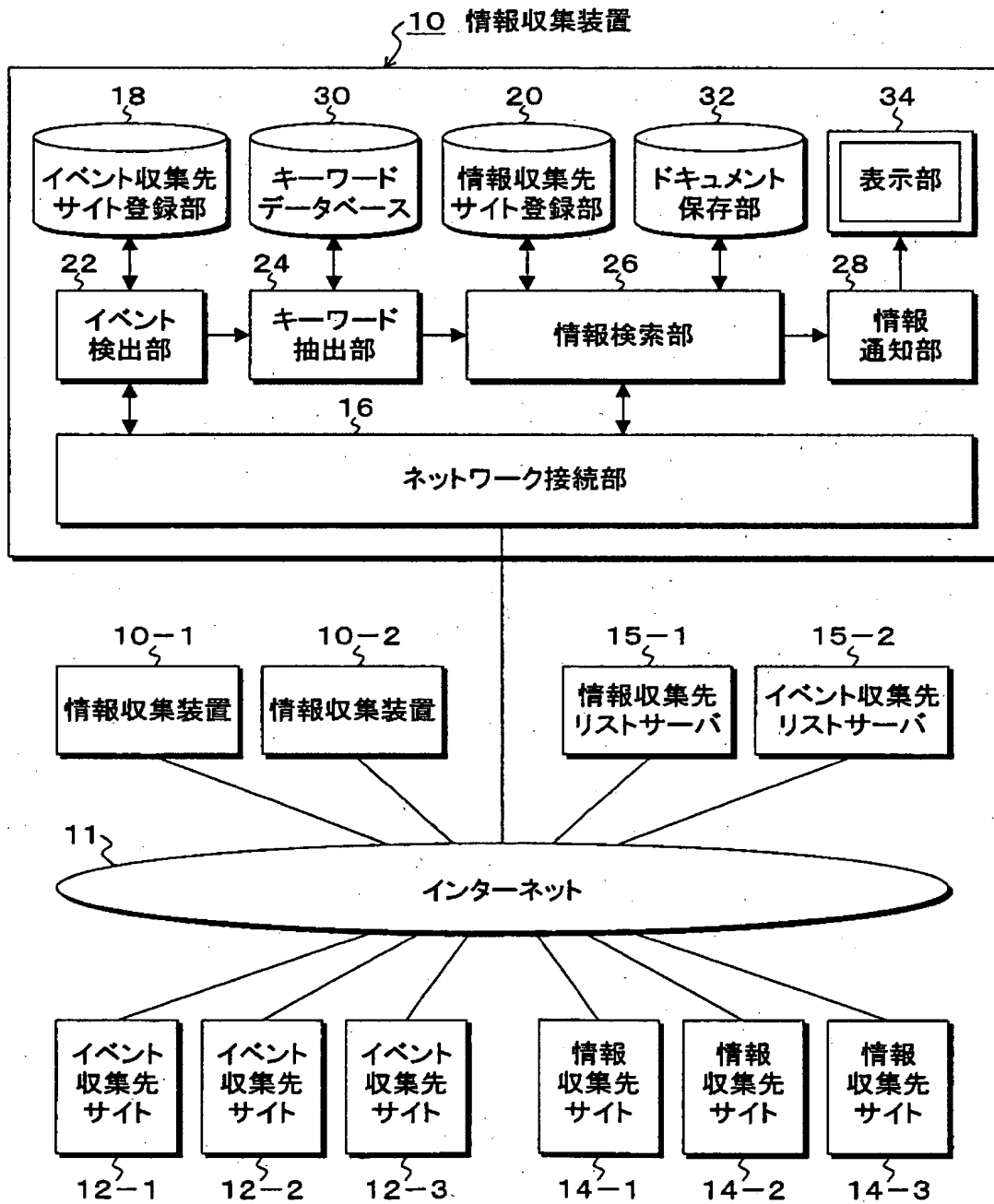
【図1】

本発明の原理構成の説明図



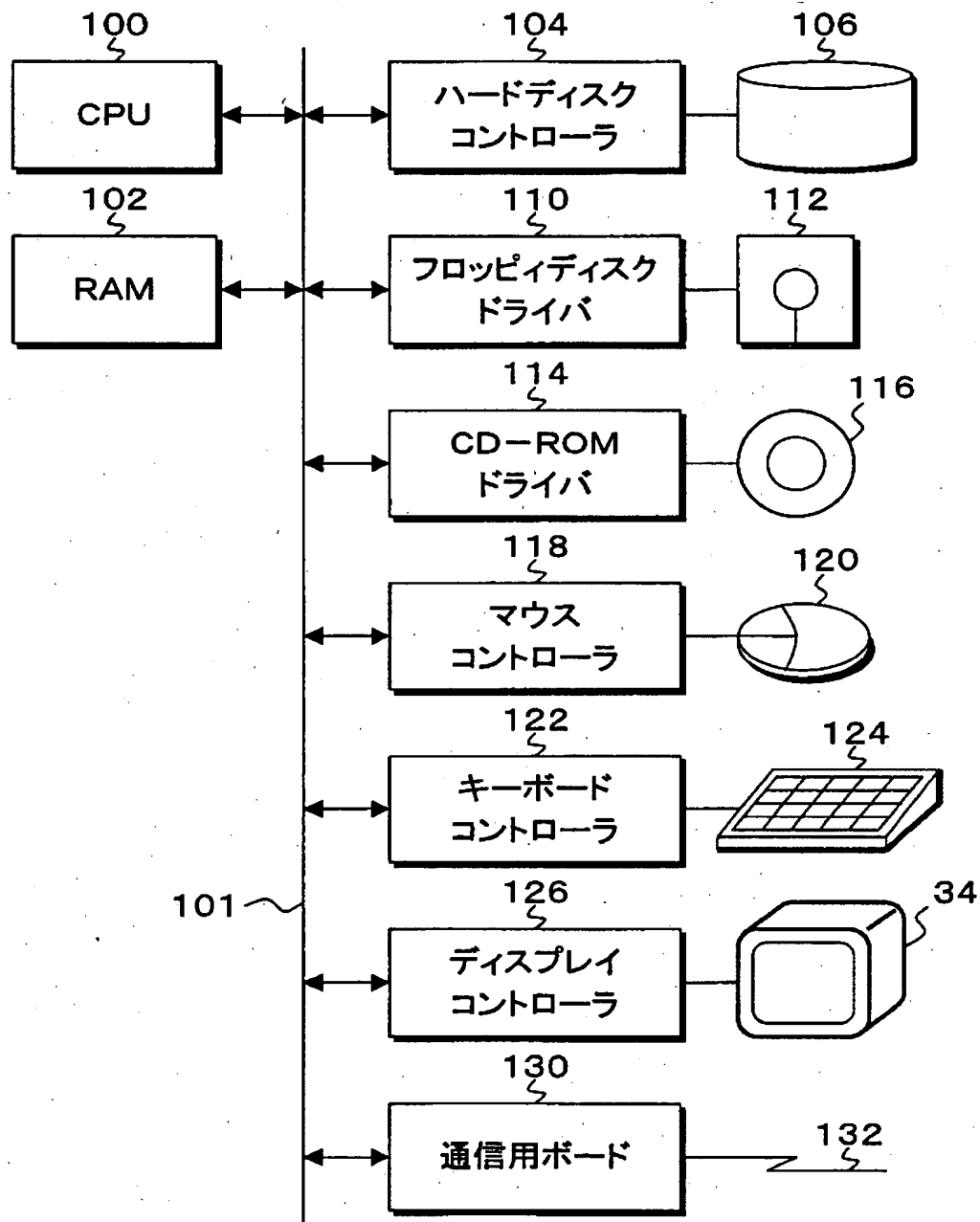
【図 2】

本発明による情報収集装置の実施形態の機能ブロック図



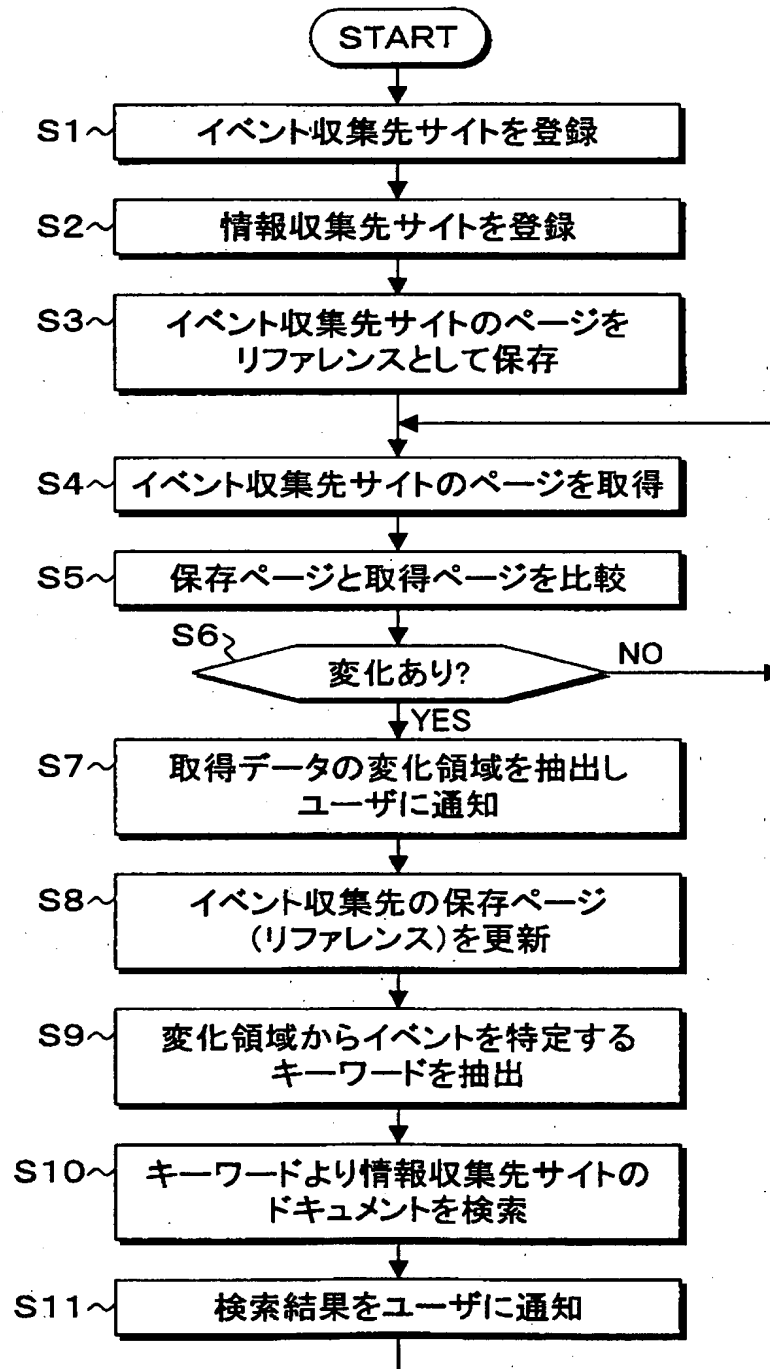
【図 3】

図2の実施形態が適用されるコンピュータのハードウェア資源の説明図



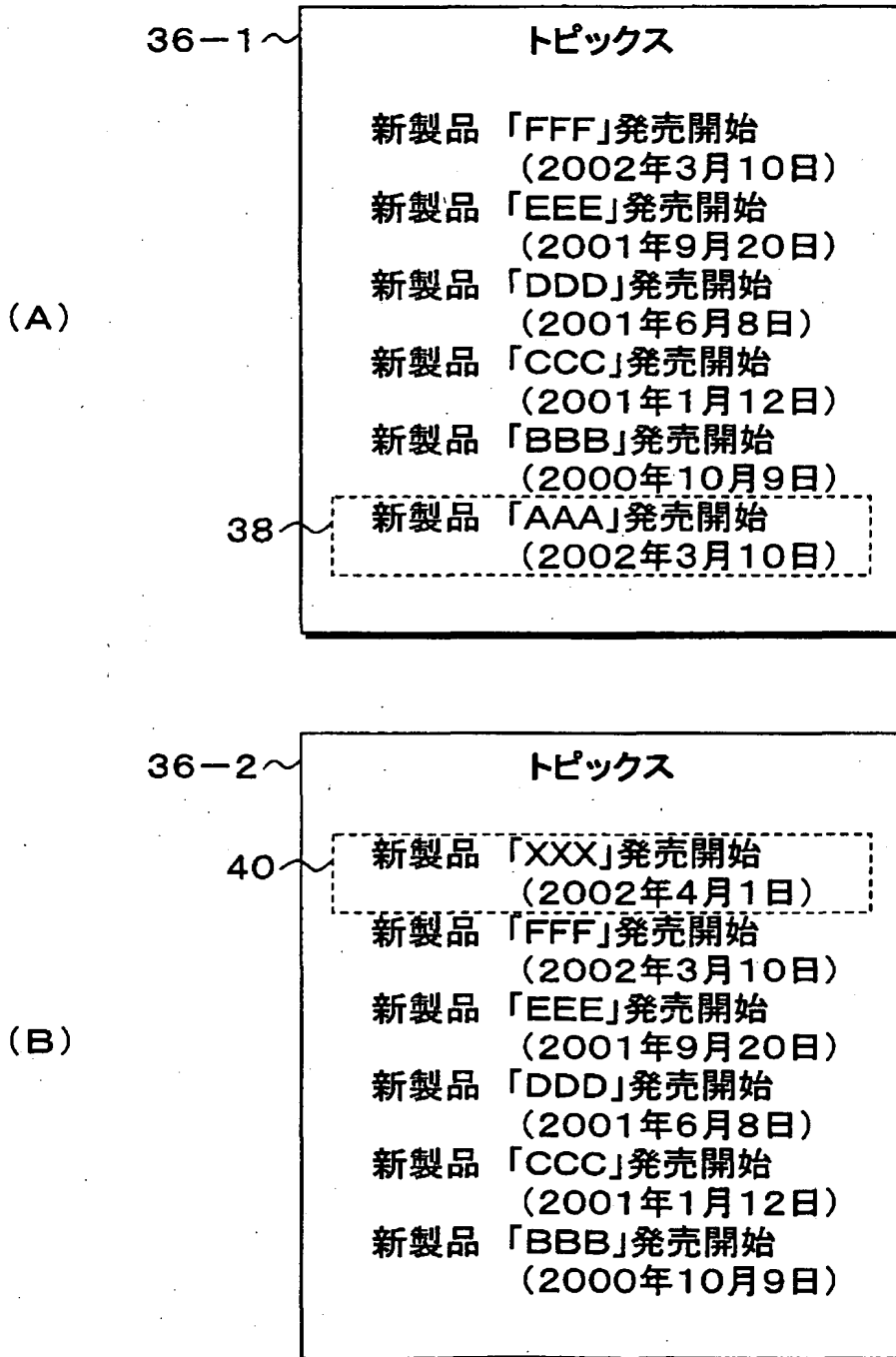
【図 4】

図2の実施形態による情報収集処理の基本的処理手順のフローチャート



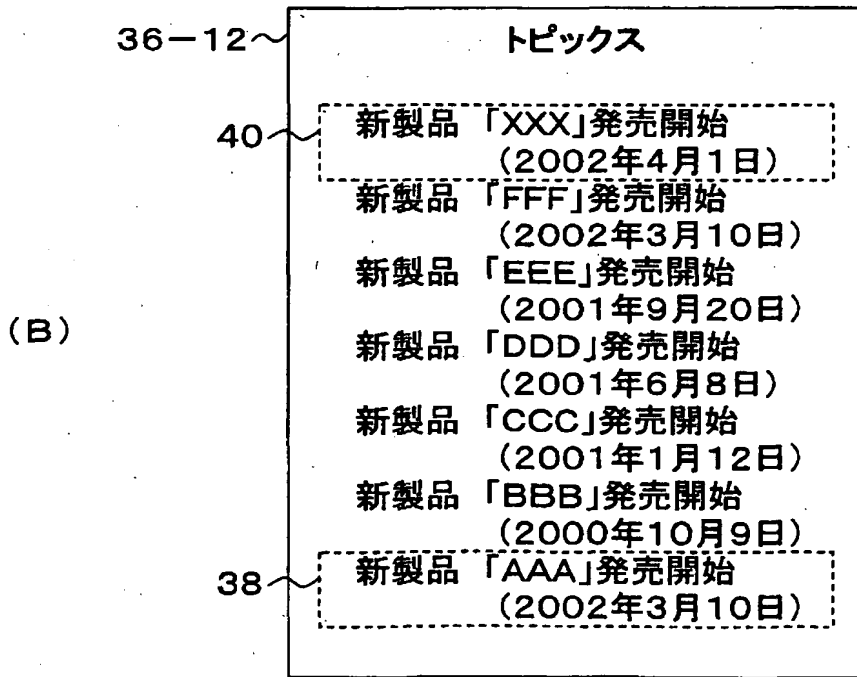
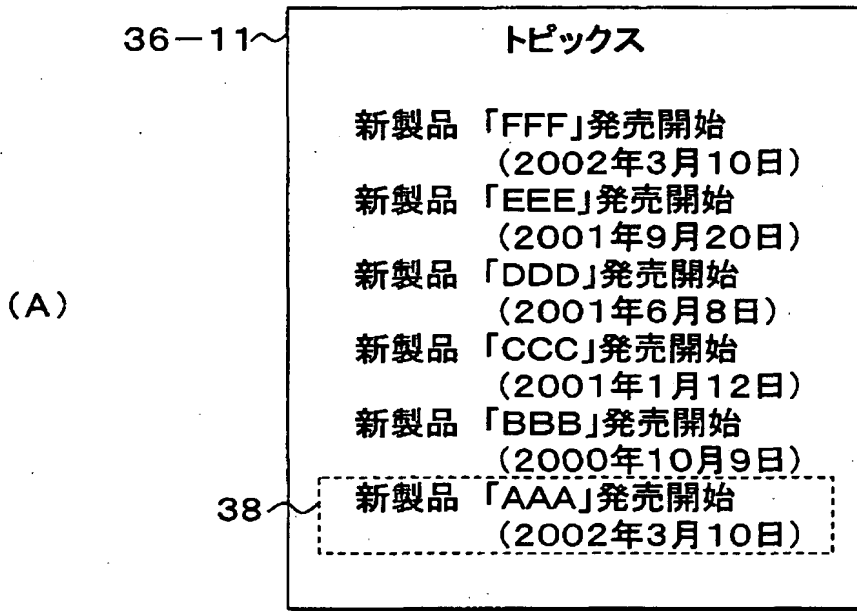
【図 5】

イベント収集先サイトから取得した新製品発売情報の説明図



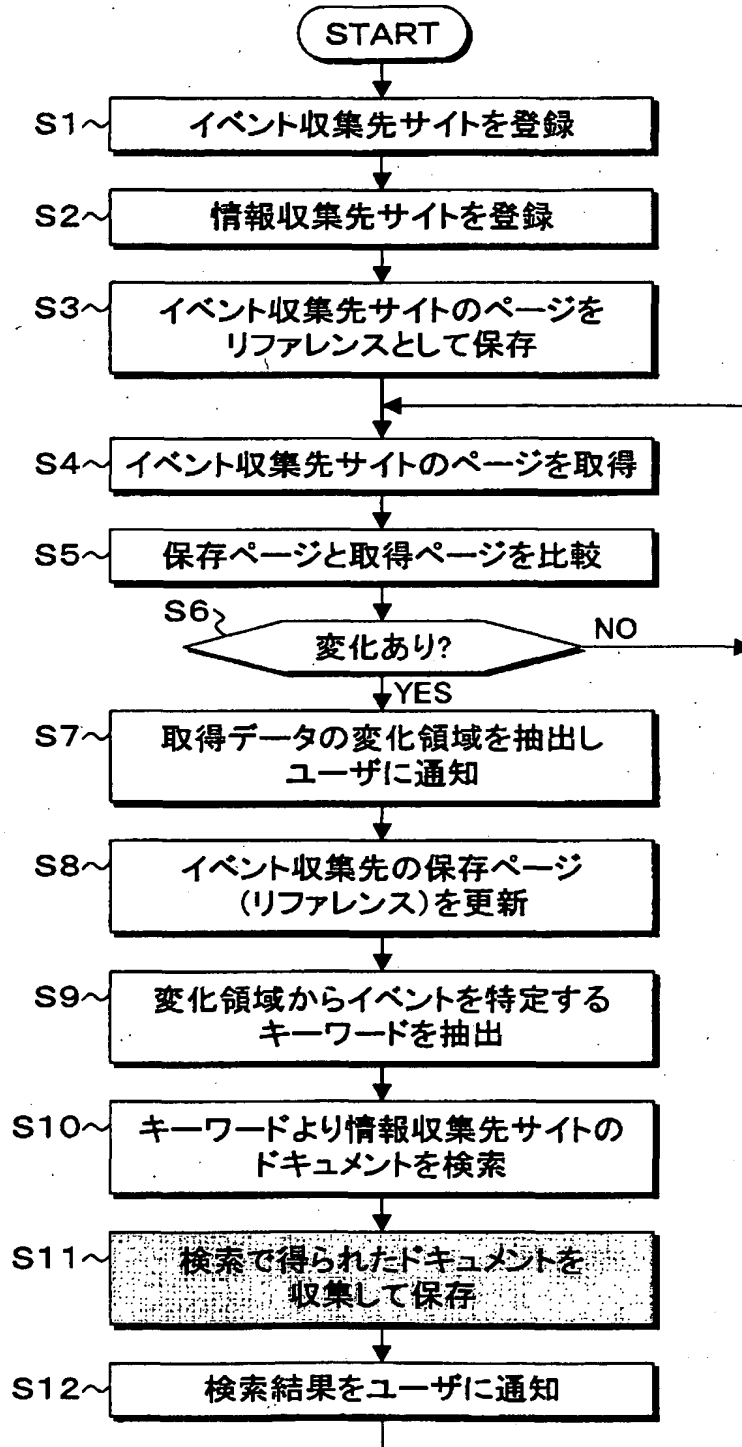
【図 6】

イベント収集先サイトから取得した新製品発売情報の他の形態の説明図



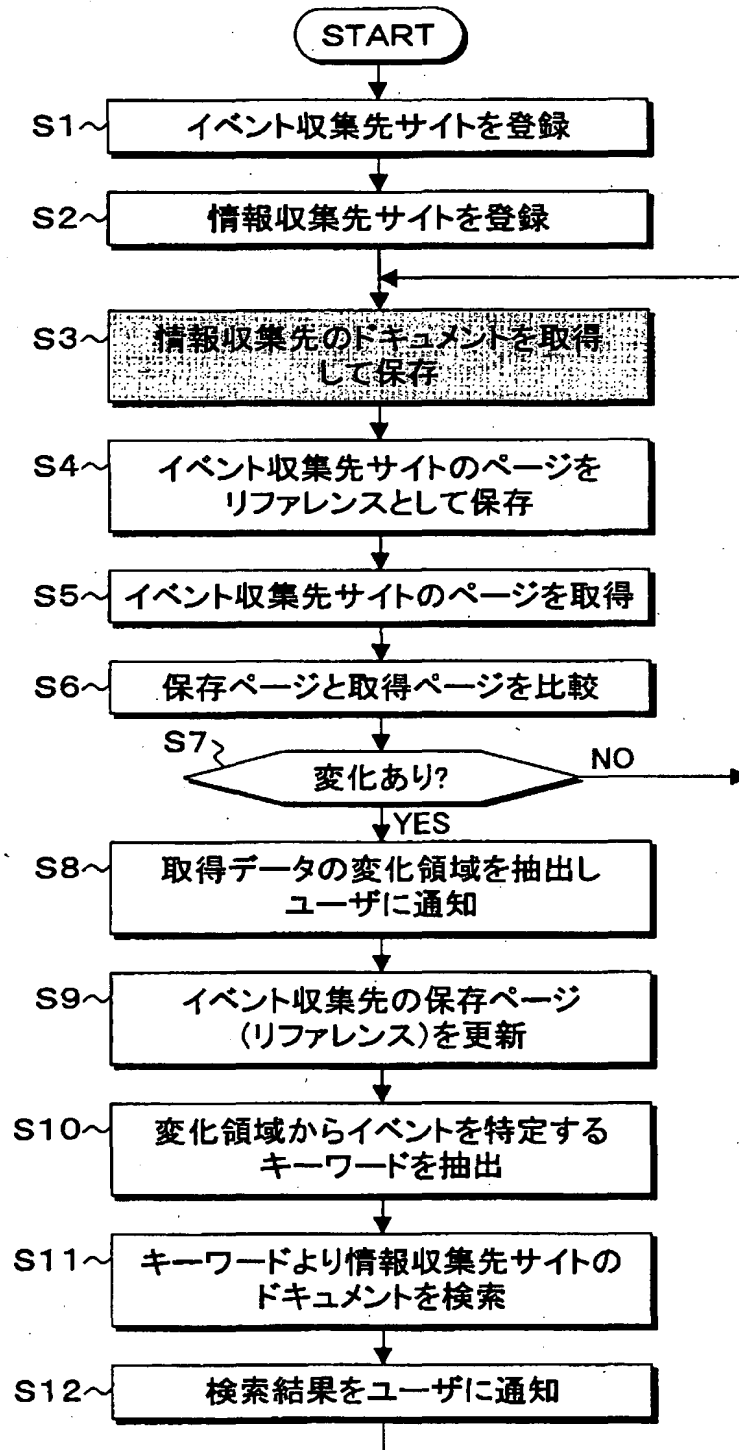
【図 7】

情報収集先サイトからキーワードにより検索したドキュメントを保存する
本発明の他の実施形態のフローチャート



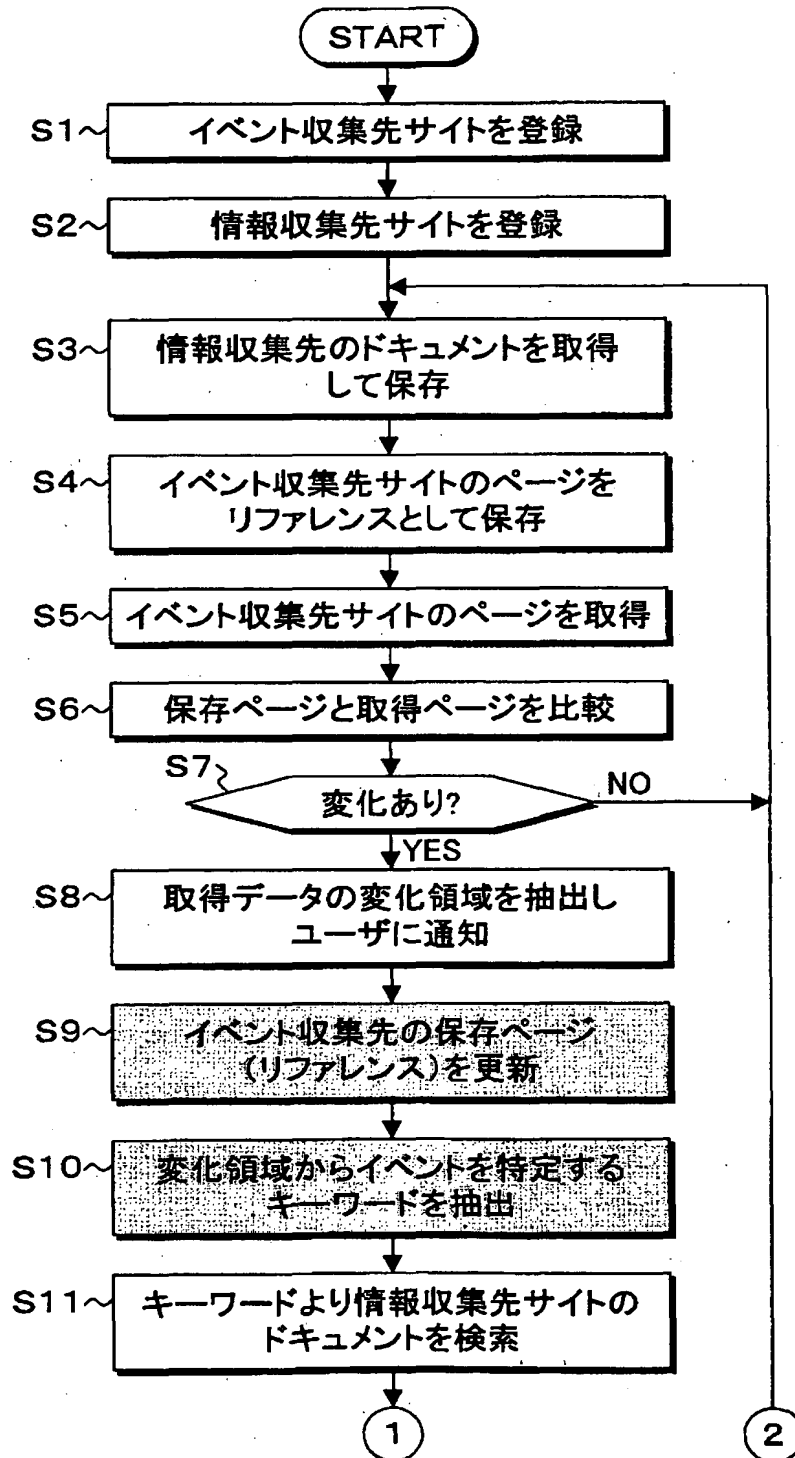
【図 8】

情報収集先サイトから収集したドキュメントを保存した後にキーワードにより保存ドキュメントを検索する本発明の他の実施形態のフローチャート



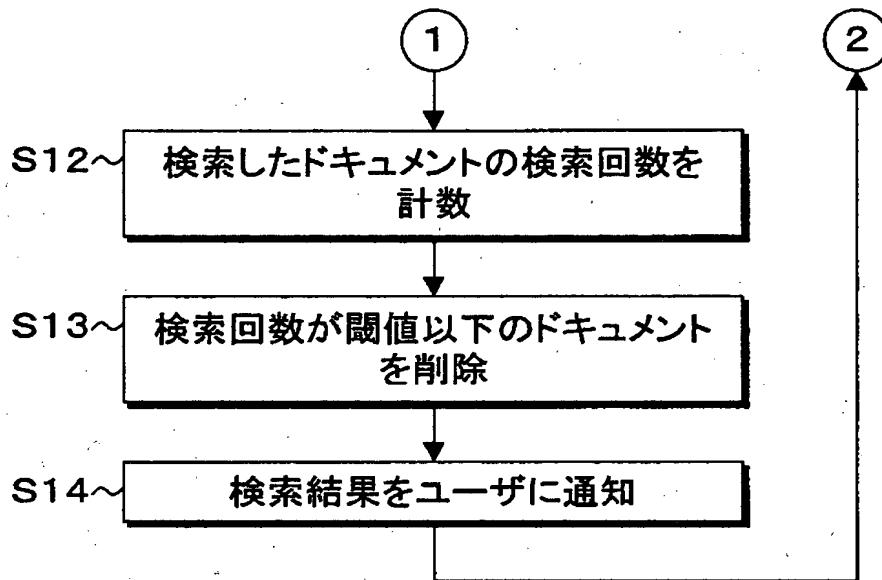
【図9】

検索回数の少ない保存ドキュメントを削除する本発明の他の実施形態のフローチャート



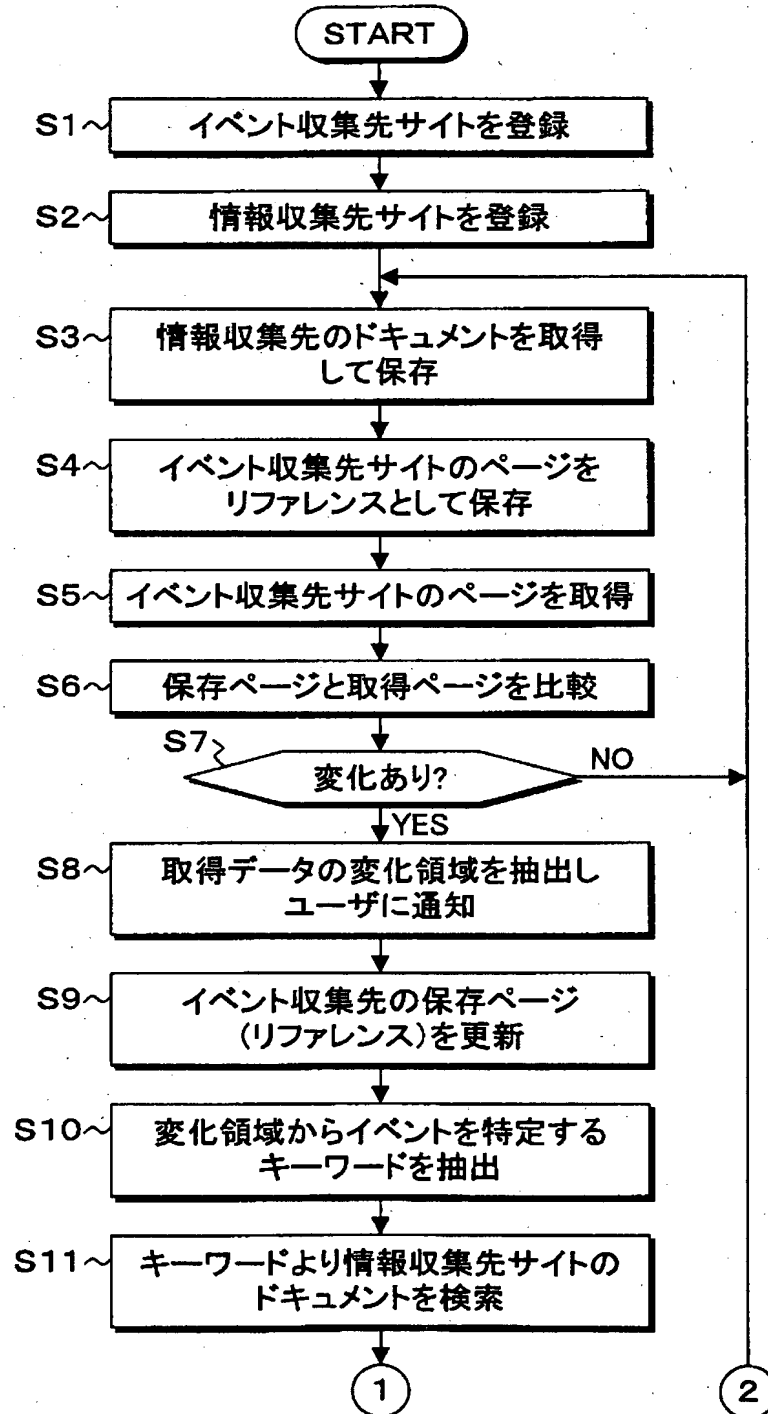
【図 1 0】

図9に続く本発明の他の実施形態のフローチャート



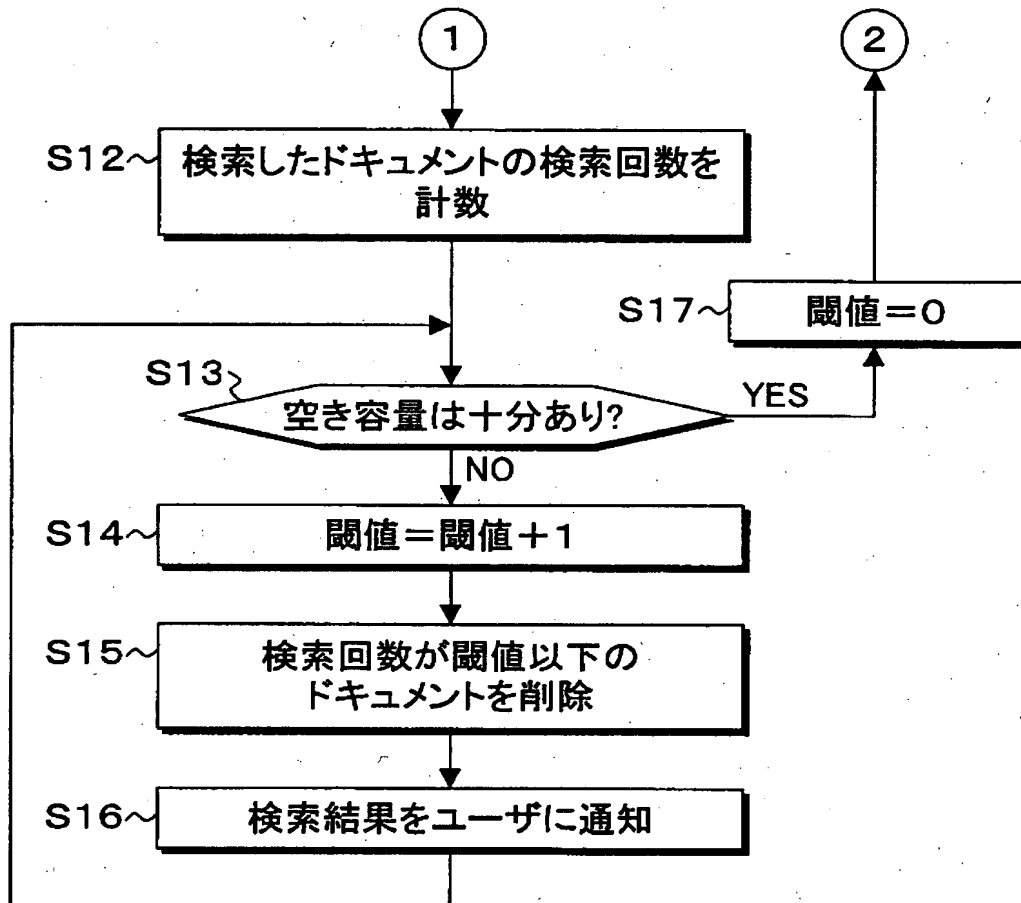
【図 1 1】

保存ドキュメントを削除する検索回数の閾値を大きくして十分な空き容量を確保する本発明の他の実施形態のフローチャート



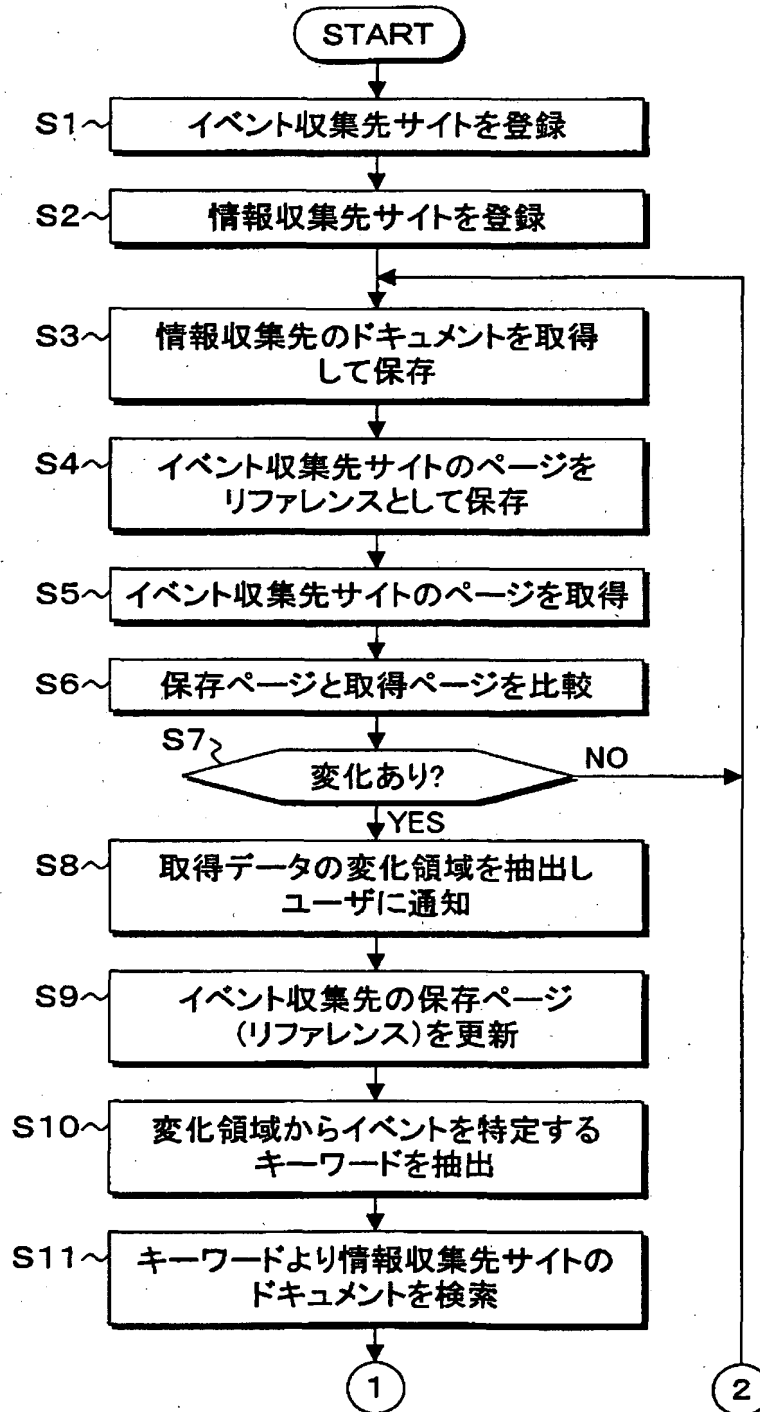
【図 12】

図11に続く本発明の他の実施形態のフローチャート



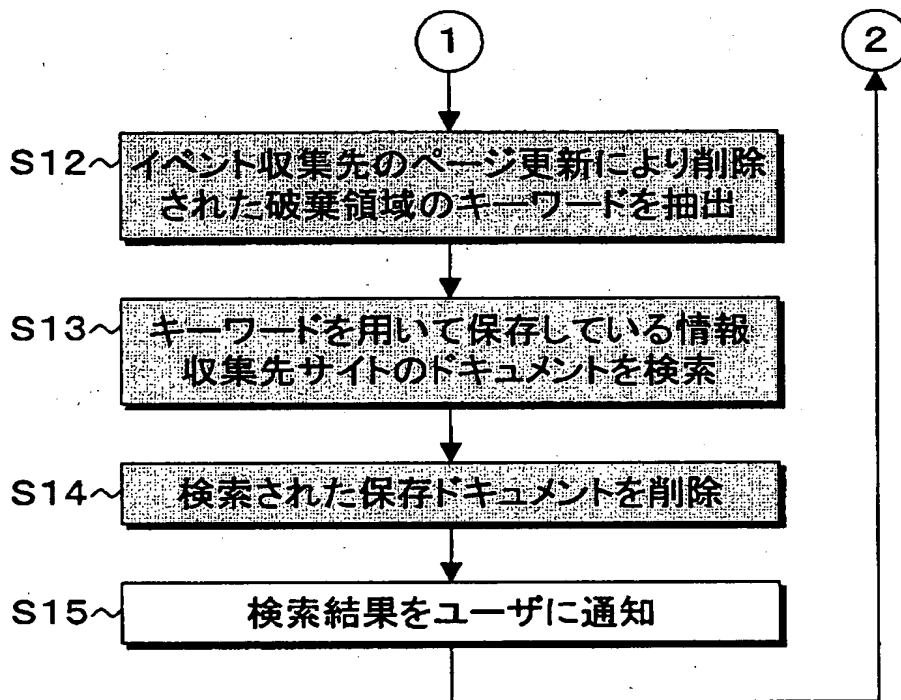
【図13】

イベント収集先サイトの情報更新で削除された破棄領域からキーワードを抽出して保存ドキュメントを削除する本発明の他の実施形態のフローチャート



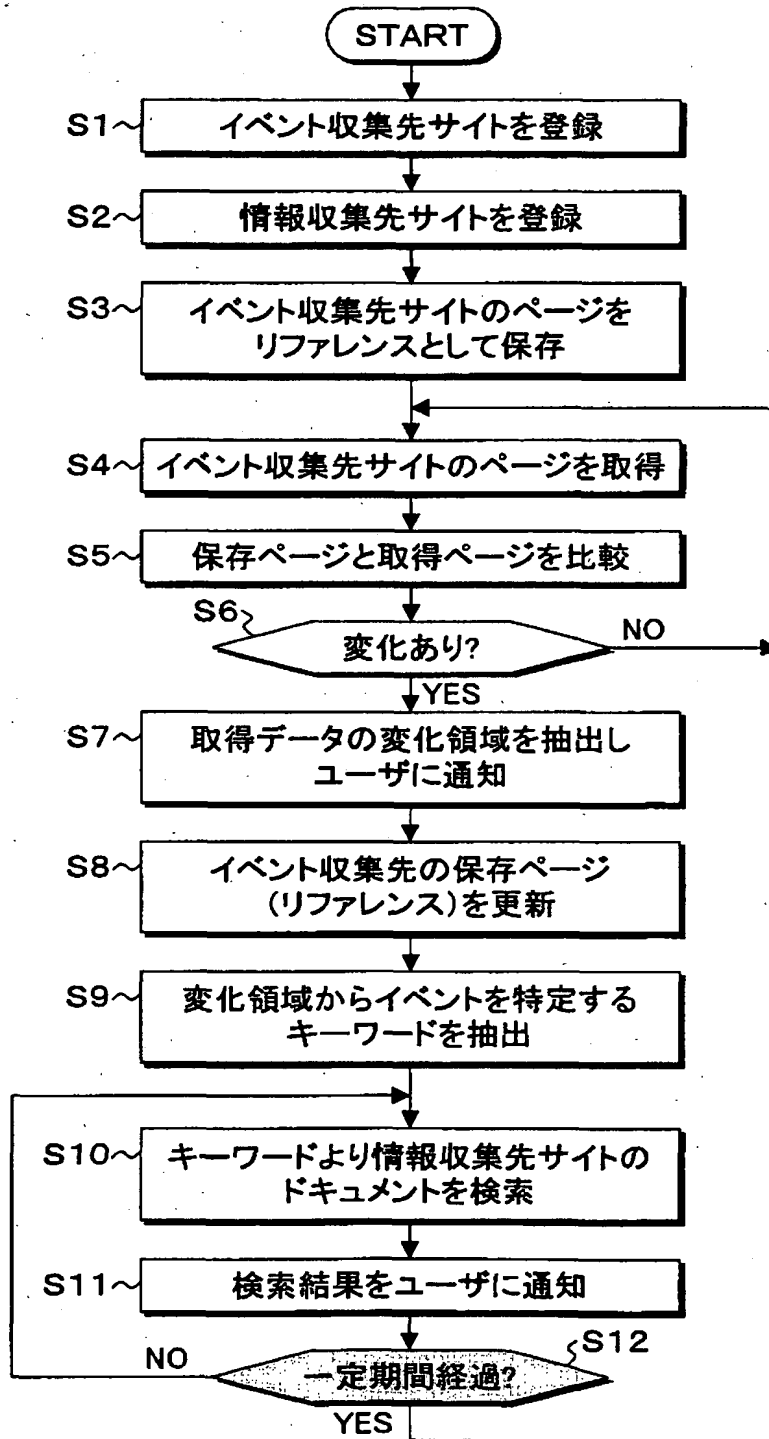
【図 1 4】

図 13 に続く本発明の他の実施形態のフローチャート



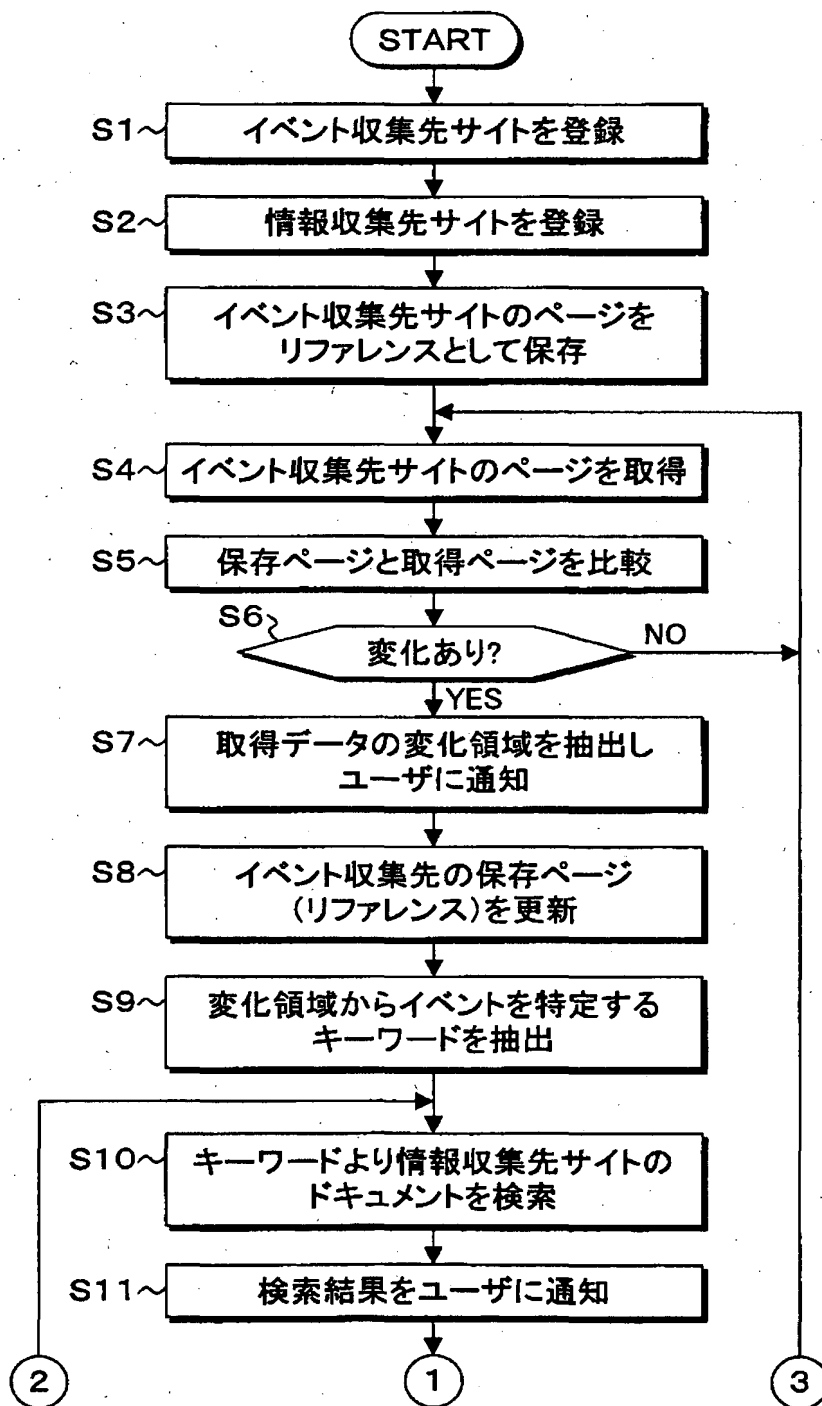
【図 15】

イベント発生検出から一定時間経過するまで定期的にキーワードによりドキュメントを検索する本発明の他の実施形態のフローチャート



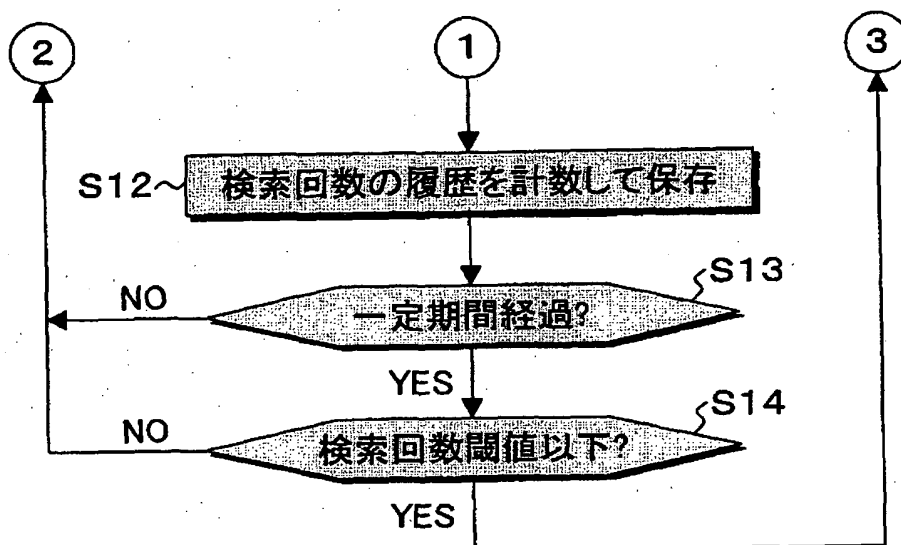
【図 1 6】

一定期間のあいだ検索回数が閾値以下であれば発生イベントの新規性が失われたとして情報収集を停止する本発明の他の実施形態のフローチャート



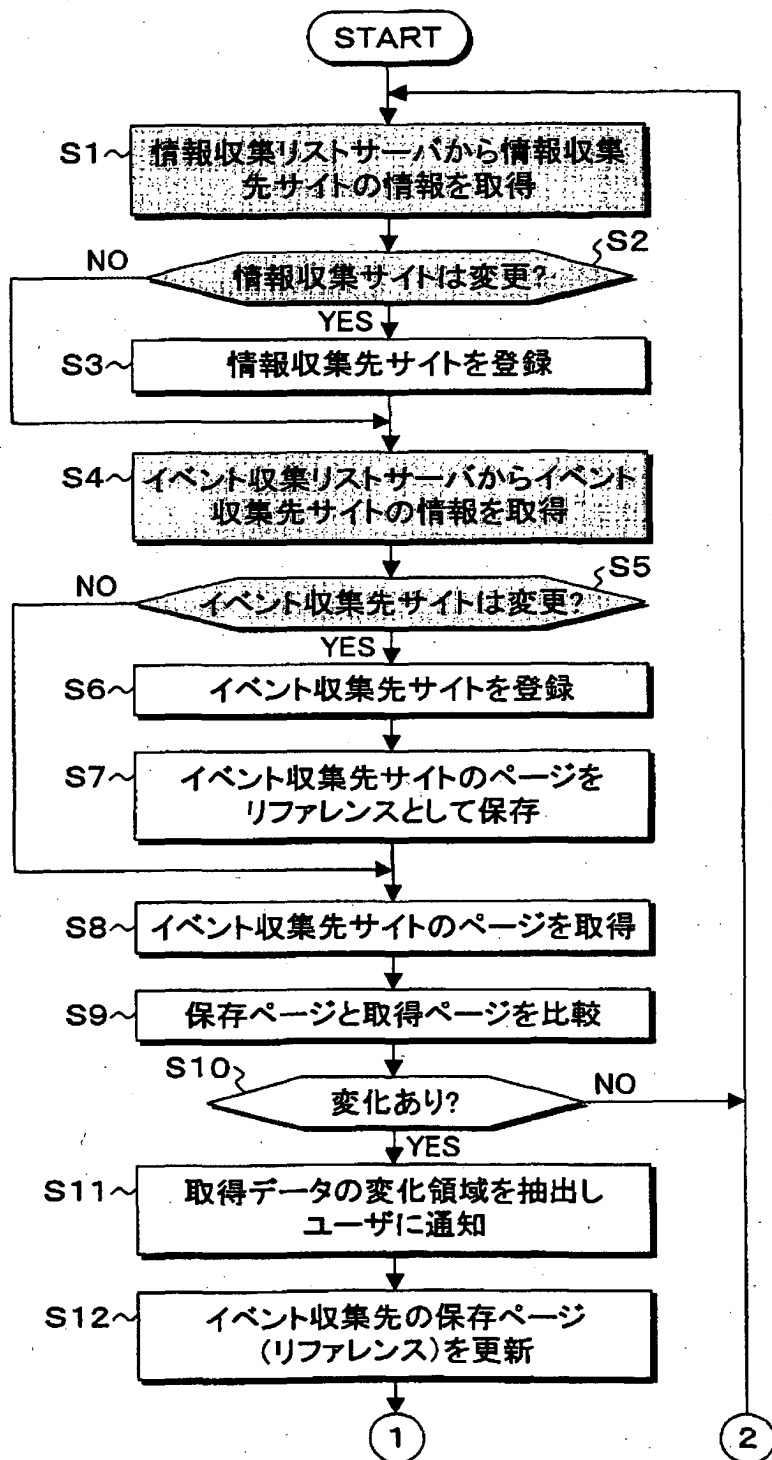
【図17】

図16に続く本発明の他の実施形態のフローチャート



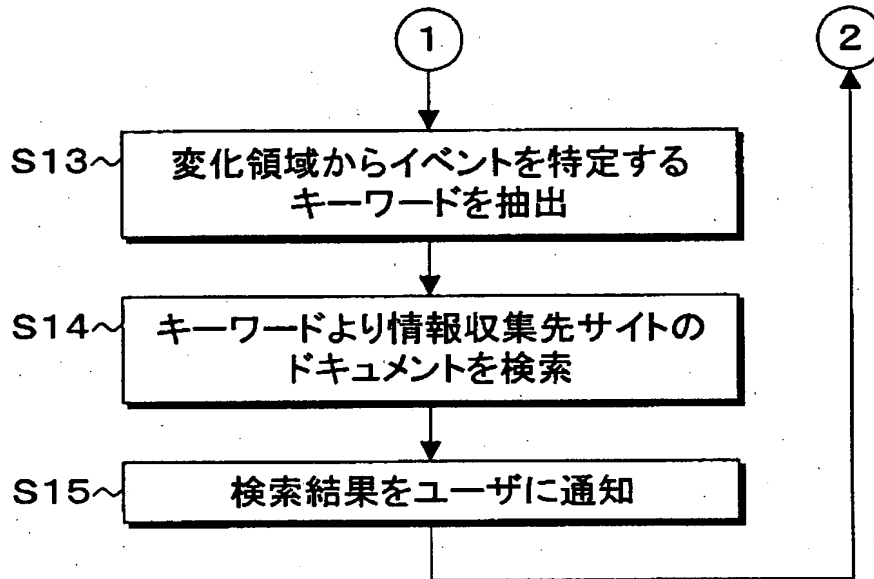
【図18】

リストサーバからイベント収集サイト及び情報収集先サイトを取得する
本発明の他の実施形態のフローチャート



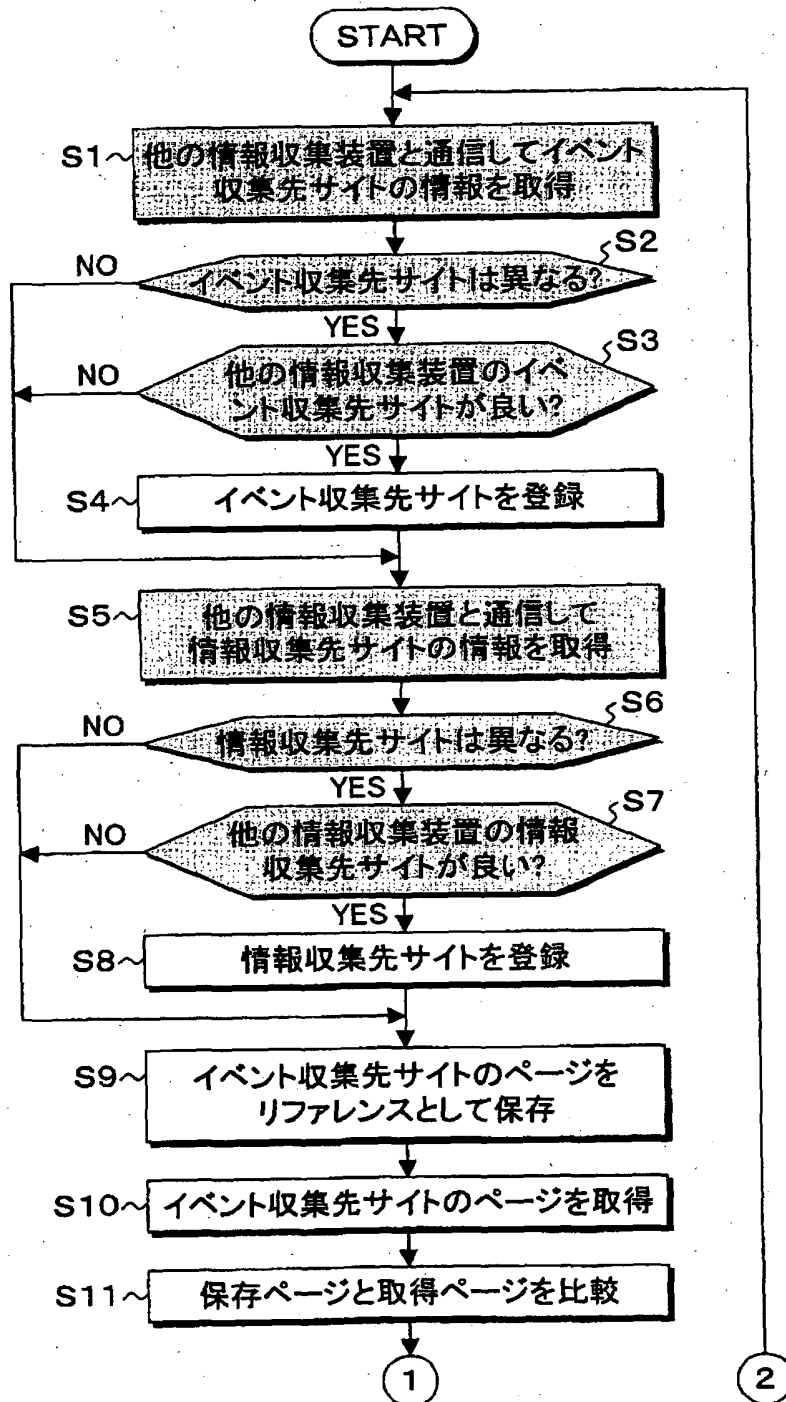
【図 1 9】

図18に続く本発明の他の実施形態のフローチャート



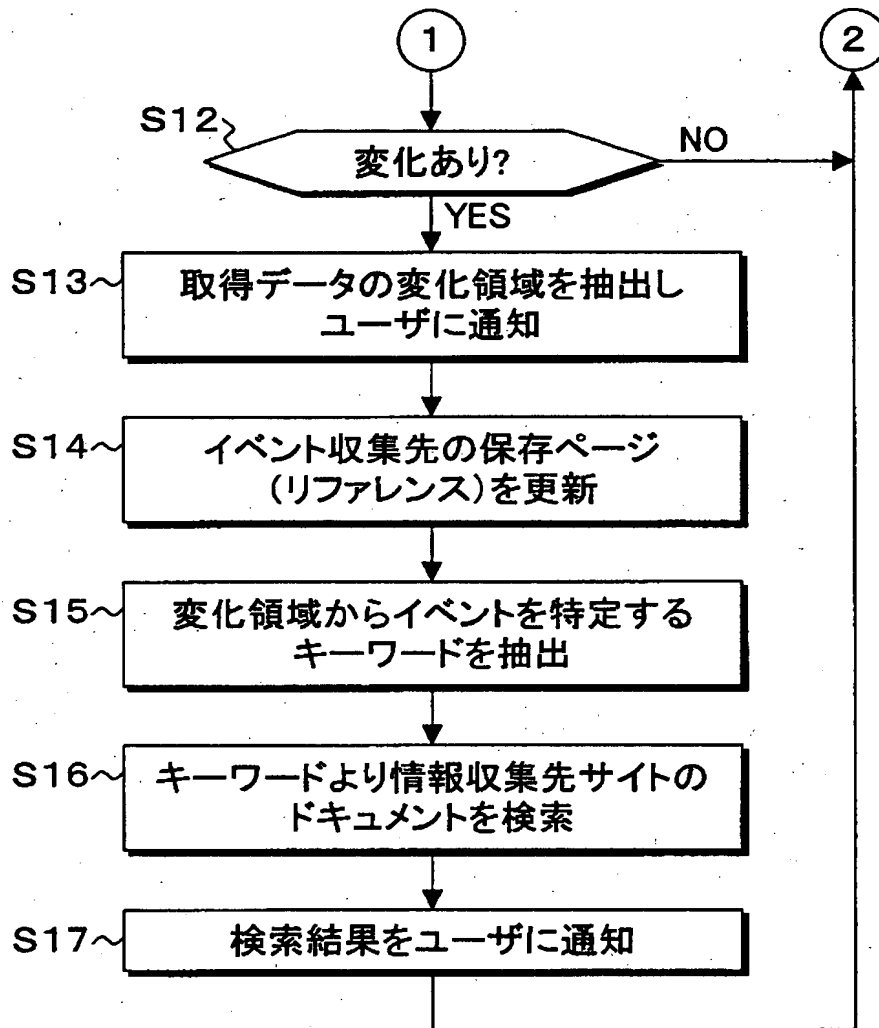
【図 20】

他の情報収集先装置からイベント収集サイト及び情報収集先サイトを取得する本発明の他の実施形態のフローチャート



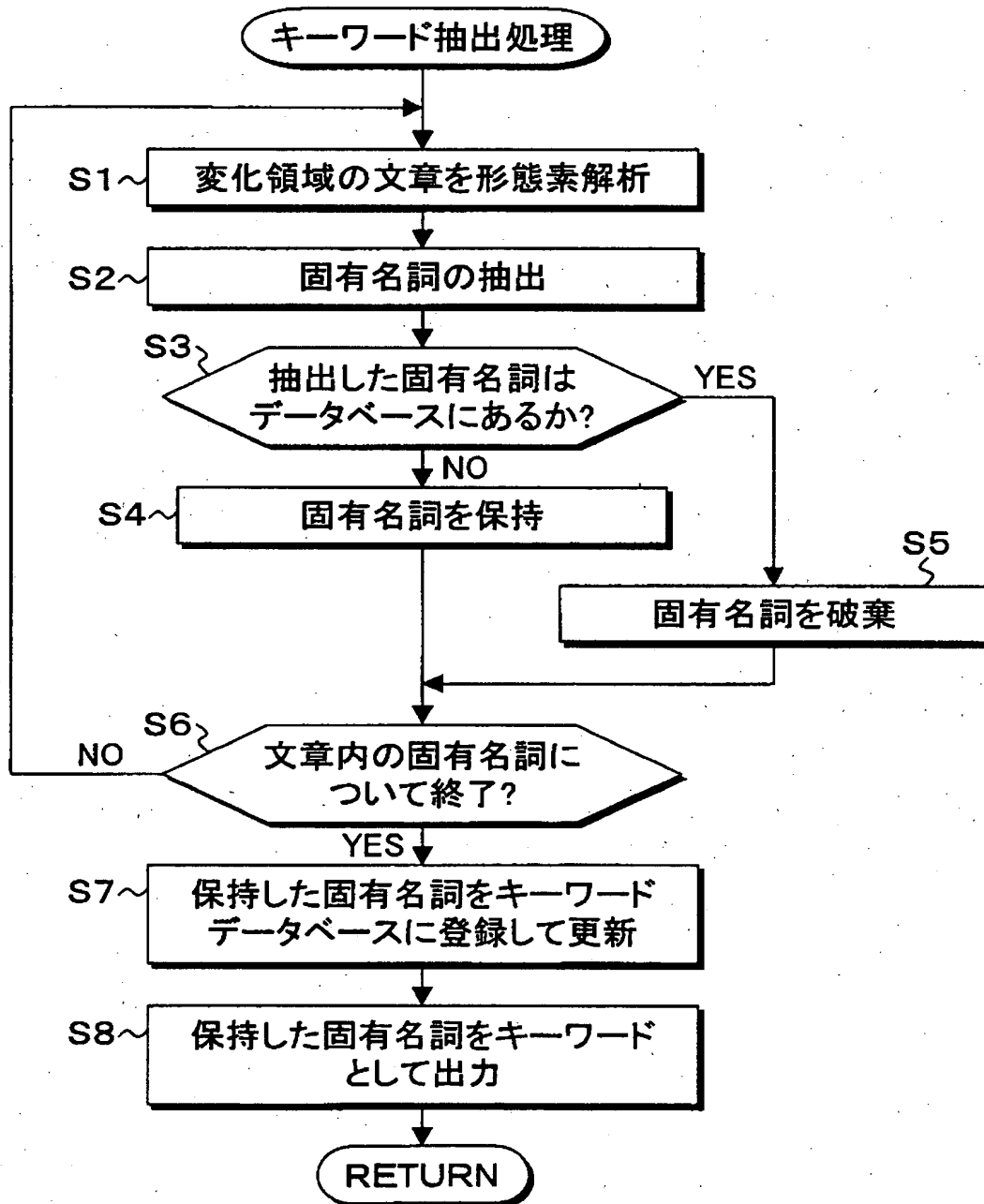
【図 2 1】

図20に続く本発明の他の実施形態のフローチャート



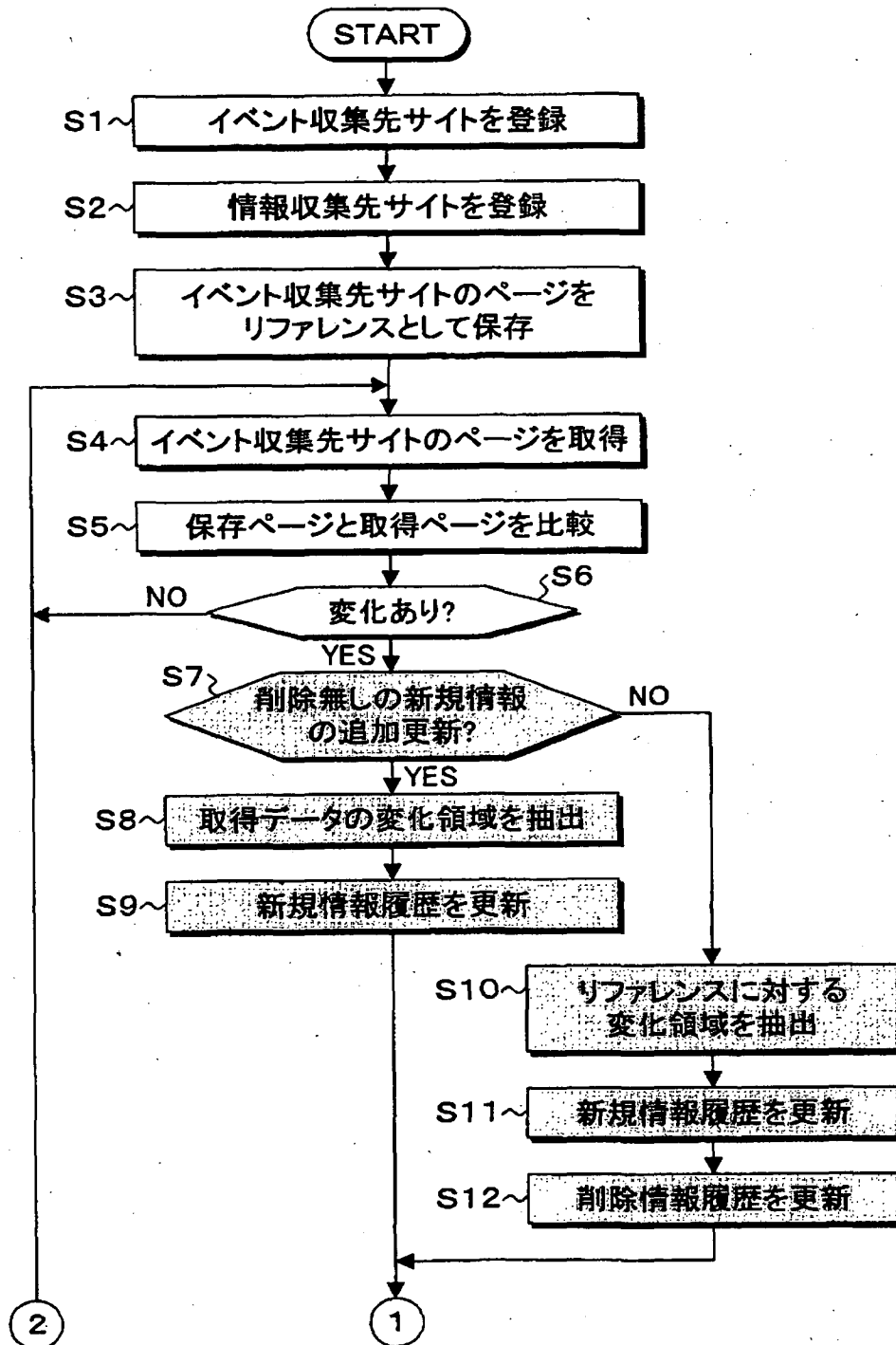
【図 2.2】

本発明におけるキーワード抽出処理のフローチャート



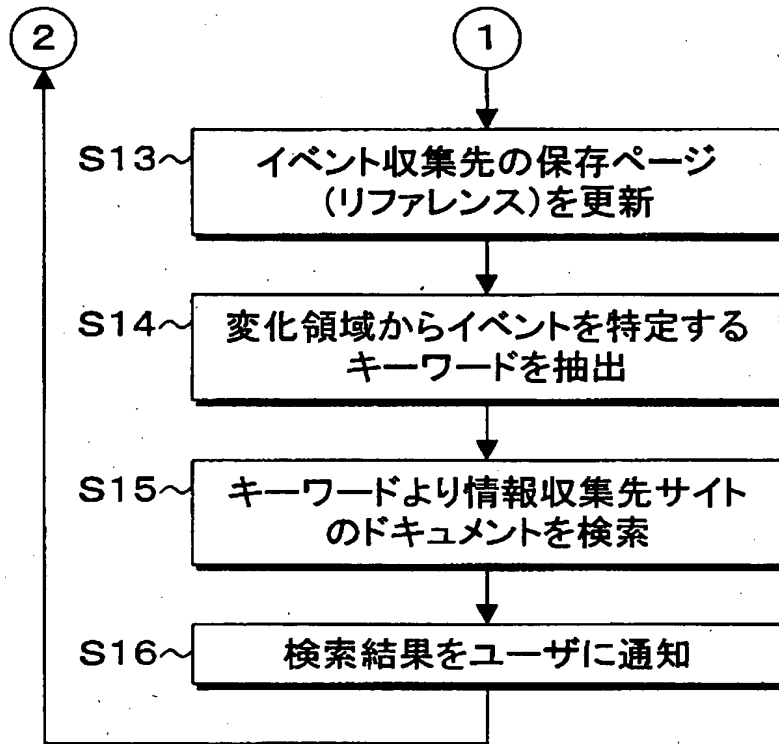
【図 23】

イベント収集先サイトの更新に伴う新規情報と削除情報の履歴を保存して利用する本発明の他の実施形態のフローチャート



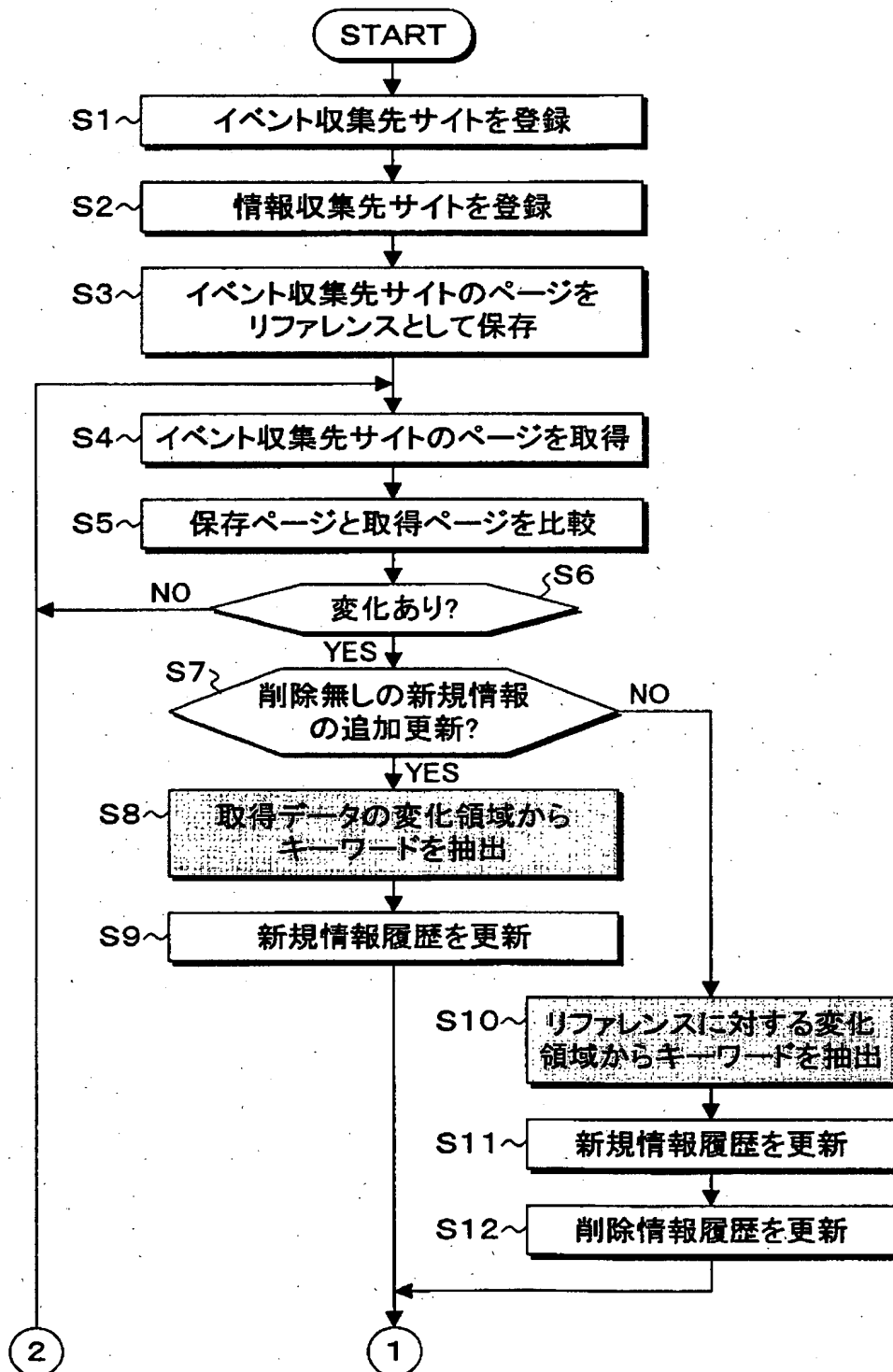
【図 2 4】

図23に続く本発明の他の実施形態のフローチャート



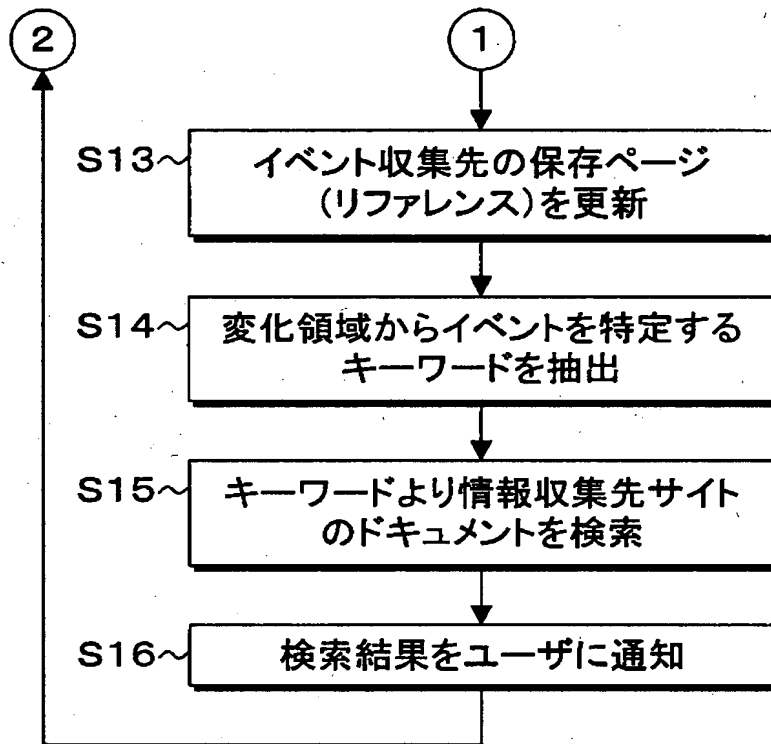
【図 25】

イベント収集先サイトの更新に伴う新規情報と削除情報の履歴をキーワードとして保存して利用する本発明の他の実施形態のフローチャート



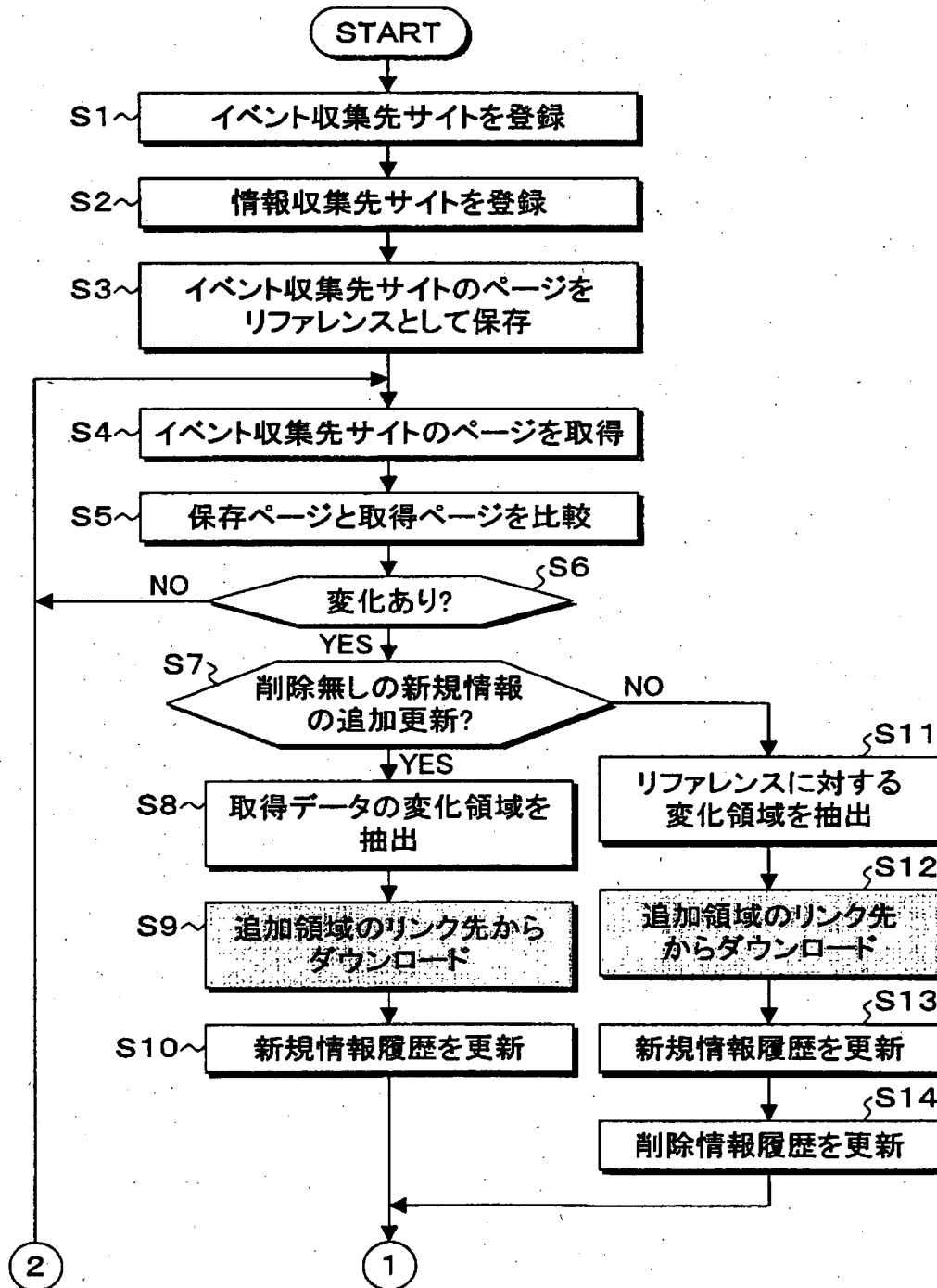
【図 2 6】

図25に続く本発明の他の実施形態のフローチャート



【図 2 7】

イベント収集先サイトの更新に伴う新規情報にある外部リンク先からドキュメントを取得して保存する本発明の他の実施形態のフローチャート



【図 2 8】

図27に続く本発明の他の実施形態のフローチャート

